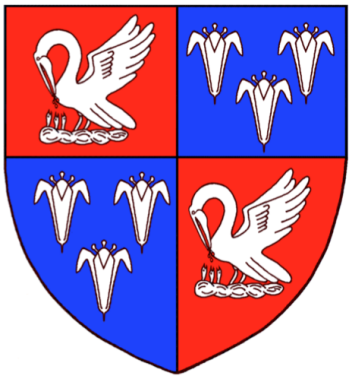
Studies on assembly and genetic variation of mitochondrial respiratory complex I



MRC Mitochondrial Biology Unit,
Corpus Christi, University of Cambridge

Polly Marino

Dissertation submitted for the degree of Doctor of Philosophy
September 2018

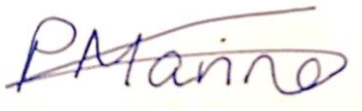
Declaration

This dissertation, entitled ‘Studies on assembly and genetic variation of mitochondrial respiratory complex I’, is the result of my own work and includes nothing that is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

This work was carried out between October 2014 and September 2018, under the supervision of Dr Judy Hirst. Information and data derived from other sources has been referenced and attributed accordingly.



Polly Marino

September 2018

Contributions

Mass spectrometry experiments were carried out by Dr Mike Harbour and Dr Shujing Ding from the MRC Mitochondrial Biology Unit, University of Cambridge.

Transcriptome data (Chapter 4) were processed by Dr Alan Robinson from the MRC Mitochondrial Biology Unit, University of Cambridge.

The human model of complex I (Chapter 6) was generated by Dan Grba from the MRC Mitochondrial Biology Unit, University of Cambridge.

Human gene variant databases (Chapter 6) were mined and collated by Dr Wei Wei, from the MRitochondrial Biology Unit, University of Cambridge.

Acknowledgments

Firstly, I would like to thank Dr Judy Hirst for the opportunity to undertake research as part of her group, and for her continued support throughout my PhD. During my time in the MRC Mitochondrial Biology Unit I have learnt many new and groundbreaking experimental techniques and acquired an array of skills that have deepened my scientific understanding, improved my critical assessment and developed my problem solving capabilities.

Thank you also to all the members of the complex I group who have made me feel so welcome and for always providing useful and insightful help – along with some unusual conversations during coffee breaks! Dr Hannah Bridges for all her guidance with cell line related work, especially her expertise on the Seahorse machine, along with Dr Andrew Jones and Dr Justin Fedor for all their advice and discussion surrounding complex I kinetics. In addition, to all the other members of the MBU, from whom I have gained a huge amount of knowledge – in particular Dr Joe Carroll, Dr Erika Fernandez-Vizarra, Dr Aurora Gómez-Duràn and Dr Mark Johnson. Jo Hutchings has provided invaluable assistance in tissue culture and always made the long hours looking after cells more enjoyable with her conversations.

To all my inspirational teachers, who always encouraged questioning and discussion – their guidance helped shape my academic path and gave me the confidence to pursue biochemistry at university and enjoy other subjects along the way. In particular to Dr Lynda Blayney, from whom I learnt so much during my placement year at the Wales Heart Research Institute, the experience was my first opportunity to undertake a research project and ultimately reinforced my desire to build on this with a PhD.

Although many things have changed since our years at school or university together my friends have always been a constant source of support, it has been brilliant to share this journey with them and follow each other's progress over recent years. I will always be grateful for their friendship and look forward to all the follows in the future.

Above all my family have always been a huge source of help and reassurance; I could not have embarked on this journey without them. My parents have provided endless encouragement and my bond with my brother and sisters has been more important than ever.

Finally, to James, for everything we have shared and all the adventures awaiting us.

To My Parents

Abstract

Complex I (NADH:ubiquinone oxidoreductase) couples electron transfer to proton translocation across the inner mitochondrial membrane, to drive the synthesis of ATP. Its distinctive L-shaped structure comprises 45 subunits, encoded by both the mitochondrial and nuclear genomes, which are assembled by a complicated modular pathway. Complex I genetic defects are the most common cause of mitochondrial disorders and often present in early childhood, with high mortality rates. Recent high-resolution electron cryo-microscopy structures of mammalian complex I provide a foundation for both interpreting biochemical and biomedical data and understanding the catalytic mechanism.

First, this thesis explores how the flavin cofactor is inserted into the NADH-binding (N-) domain of complex I. Genetic manipulation of cultured human cells, to starve them of flavin, revealed a hierarchal impact on the mitochondrial flavoproteome. High riboflavin content in the growth media ameliorated observed phenotypes, requiring cell conditioning in low riboflavin conditions. CRISPR knockout of the putative mitochondrial flavin transporter SLC25A32 demonstrated the severe impact of decreased flavin on complexes I and II, and mass spectrometry 'complexome' analyses suggest that the N-domain is still assembled onto complex I in the absence of the flavin.

Second, the model organism *Yarrowia lipolytica* was used to assess the importance of residues in the quinone-binding site of complex I. Three residues with proposed roles in binding the quinone head-group were targeted. One variant was catalytically inactive, while two retained some activity. They showed decreased ability to reduce physiologically-relevant, long chain quinones, but their ability to reduce short-chain analogues was affected less severely. The results suggest a complicated picture in which interactions between the protein and both the hydrophilic quinone head-group and hydrophobic isoprenoid chain contribute to quinone-binding affinity and catalysis.

Finally, a model for human complex I, generated from a recent high-resolution structure of mouse complex I, was used to investigate whether the pathogenicity of human variants could be predicted. Structural information on variant residues, including their secondary structure, proximity to key features and surface exposure, was collated and the power of each property to predict pathogenicity investigated. The analysis was then extended to the whole structure, to identify potential pathogenic hotspots in the enzyme, inform future studies of functionally important regions in complex I, and aid the diagnosis of clinically relevant pathogenic variants.

Abbreviations

Additionally, standard chemical symbols, SI units and the one and three letter abbreviations for the 20 biologically relevant amino acids are used throughout. The letters A, C, G, T, and U represent the nucleotides adenosine, cytidine, guanosine, thymidine and uridine, respectively. The nomenclature of complex I subunits differs across species, therefore the human names for the subunits are used in this thesis.

$\Delta\Psi$	membrane potential
Δp	proton motive force
143B CyWT	immortalised human osterocarcoma patient cybrid cells
ACN	acetonitrile
ADP	adenosine diphosphate
$APAD^+$	3-acetlypyridine adenine dinucleotide
AOX	alternative oxidase
ATP	adenosine triphosphate
BCA	bicinchronic acid
BLAST	basic local alignment search tool
BN-PAGE	blue native polyacrylamide gel electrophoresis
CHAPS	3-[(3-cholamidopropyl)-dimethylamminio]-1-propane sulphonate
CRISPR	clustered regularly spaced short palindromic repeats
Cryo-EM	cryogenic electron microscopy
Cyt_c	cytochrome *c*
DDM	n-dodecyl-β-D-maltoside
DMSO	dimethyl sulphoxide
dNADH	deamino reduced nicotinamide adenine dinucleotide
DQ	decylubiquinone
DRP1	dynamin-1-like protein
DTT	dithioreitol
ECAR	extracellular acidification rate

ECL	electrochemiluminescence
EDTA	ethylenediaminetetraacetic acid
EPR	electron paramagnetic resonance
ETC	electron transfer chain
FAD	flavin adenine dinucleotide
FADS	flavin adenine dinucleotide synthase
FADS1	flavin adenine dinucleotide synthase isoform 1 (mitochondrial)
FCCP	carbonyl cyanide p-trifluoromethoxyphenylhydrazone
FeCN	potassium hexacyanoferrate
FeS	iron-sulphur
FLAD1	flavin adenine dinucleotide synthase gene
FMN	flavin mononucleotide
gDNA	genomic DNA
HAP1	male chronic myelogenous leukemia cell line KBM-7
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HGMD	human gene mutation database
HPLC	high-pressure liquid chromatography
HRP	horseradish peroxidase
IMM	inner mitochondrial membrane
IMS	intermembrane space
LHON	Leber's hereditary optic neuropathy
MOPS	3-(N-morpholino)propanesulphonic acid
MSH buffer	mannitol, sucrose, 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid buffer
mtDNA	mitochondrial deoxyribonucleic acid
MWCO	molecular weight cut-off
NAD^+	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NBT	nitroblue tetrazolium
NCBI	US National Centre for Biotechnology Information

nDNA	nuclear deoxyribonucleic acid
NDH2	alternative NADH dehydrogenase
OCR	oxygen consumption rate
OMM	outer mitochondrial membrane
OXPHOS	oxidative phosphorylation
PAM	protospacer motif/point accepted mutation
PCR	polymerase chain reaction
PDB	protein databank
PL	proteoliposome
PMSF	phenylmethanesulphonylfluoride
PVDF	polyvinylidene fluoride
Q	ubiquinone
Q_1	coenzyme quionone-1
Q_6	coenzyme quionone-6
Q_9	coenzyme quionone-9
Q_{10}	coenzyme quionone-10
QH_2	ubiquinol
Q-site	ubiquinone binding site
RFK	riboflavin kinase
RFK KD	riboflavin kinase knockdown
ROS	reactive oxygen species
RNA	ribonucleic acid
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNA	single guide ribonucleic acid
shRNA	short hairpin ribonucleic acid
SILAC	stable isotope labelling of amino acids in cell culture
siRNA	small interfering ribonucleic acid
SLC25A32	putative mitochondrial folate/flavin carrier
SLCKO	putative mitochondrial folate/flavin carrier knockout

SOD	superoxide dismutase
TMH	transmembrane helix
TMRM	tetramethylrhodamine
VDAC1	voltage dependent ion channel 1
WT	wild type
YPD	yeast extract peptone dextrose

Table of Contents

Chapter 1

Introduction

1.1. Energy transduction in biological systems

Biological systems must harness energy from the environment in order to drive thermodynamically unfavourable reactions. They achieve this through energy transducing membranes and the generation of ATP from ADP + P_i. The plasma membrane of prokaryotes, thylakoid membrane of chloroplasts and the inner membrane of mitochondria all act in this way. The fundamental components of energy transducing membranes are: an ion impermeable barrier containing protein complexes capable of pumping protons, and the movement of electrons along a chain of proteins with increasing redox potential to release energy, which can be harnessed to move protons across the membrane against their concentration gradient. In mitochondria the accumulation of protons in the intermembrane space forms the proton motive force (Δp), which comprises two elements: the ΔΨ due to the charge imbalance and ΔpH from the difference in proton concentration. Therefore, these membranes require two types of proton pumps – those that utilise the Gibbs free energy from electron transfer to move protons across the membrane to build Δp and those that couple the release of Δp to energy demanding processes, such as the synthesis of ATP (Nicholls & Ferguson, 2013). In mitochondrial membranes the two types are the electron transport chain (ETC), where a series of protein complexes couple electron transfer to the translocation of protons, and ATP synthase (complex V), which uses the controlled collapse of Δp to drive ATP synthesis.

The operation of energy transducing membranes was first demonstrated by Peter Mitchell in his groundbreaking work on the chemiosmotic process in 1961, although it received much criticism at the time (Mitchell, 1961). The prevailing hypothesis was that energy was stored as a high-energy intermediate, despite little evidence of this species. Increased understanding of proton pumping by the complexes of the ETC gave further weight to the chemiosmotic theory and Mitchell received the 1978 Noble Prize in Chemistry for his work.

1.2. Mitochondria

Mitochondria are the energy transforming organelles found in most eukaryotes, although they also perform a number of other fundamental roles. It is believed that these organelles evolved as described by the endosymbiotic theory, first postulated by Margulis in 1967 (Sagan, 1967). Driven by the evolution of the Earth, from an oxygen free to oxygen rich environment, it was hypothesised that an aerobe was ingested by an anaerobic prokaryote resulting in a mutually beneficial relationship. The engulfed cell would become a double-walled organelle, the inner lipid bilayer from the original plasma

membrane surrounded by an additional bilayer formed when ingested into a vesicle (Sagan, 1967). The endosymbiont theory has further been substantiated with the publication of the genome sequence of *Rickettsia prowazekii*; an ancestor of this parasite was thought to be the engulfed species (Andersson *et al*., 1998). With evolutionary time the symbiotic relationship has developed into cells with a highly specialised energy transducing organelle, which has retained a limited genomic DNA content and the ability to independently replicate (Nunnari & Suomalainen, 2012). Through transfer of genomic material to the nucleus the circular mtDNA of mammals now contains genes for only 13 proteins, 22 tRNAs and 2 mitochondrial ribosomal subunits.

Mitochondria have developed a highly adapted structure, with a largely permeable outer membrane (OMM) and an impermeable inner membrane (IMM), across which Δp is maintained. The IMM also forms many invaginations, known as cristae, projecting into the protein rich matrix, Figure 1.1. The first electron micrographs characterised the intricate structure of mitochondria, which is common to many different tissues, highlighting its importance to the function of the organelle (Palade, 1952). Since then imaging techniques have further informed knowledge of the mitochondrial structure, such that the IMM is considered to have two subdomains, the inner boundary membrane (IBM) and cristae membrane (CM). The IBM forms contact sites with the OMM, containing the import TOM and TIM machinery along with the ADP/ATP carrier responsible for ATP export. The tight association of the two membranes is also necessary for their coordinated fusion and fission (Vogel *et al*., 2006). In contrast the CM, comprising most of the IMM, forms invaginations where parts of the membrane are juxtaposed to one another and between the two subdomains are the narrowings of the cristae junctions, which may form barriers between the intermembrane space (IMS) and intracristal space.

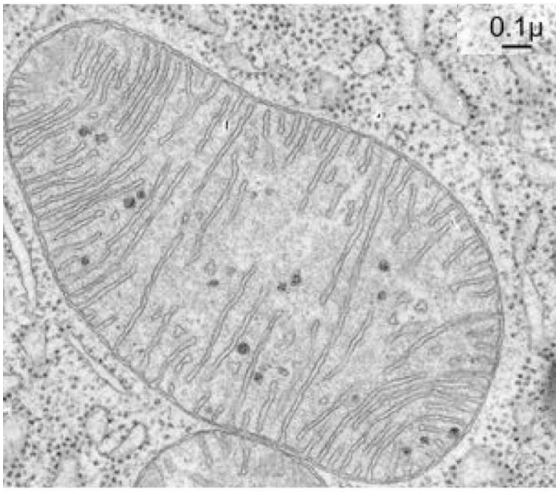


Figure 1.1 Electron micrograph of a mitochondrion from bat pancreas. The image shows the defining features of mitochondria including a double membrane. The inner membrane invaginations – cristae – are clearly visible extending into the mitochondrial matrix. Dark points within the matrix are mitochondrial ribosomes. Image taken from Porter & Bonneville, 1973.

Although the complex structure of individual mitochondria has long been appreciated, the importance of their dynamics, through fission and fusion events, was not fully understood, with the general paradigm of isolated mitochondria rather than an interacting network. However, live cell imaging showed that mitochondria exist as extended reticular networks contacting the nucleus, Golgi and cytoskeleton. The equilibrium between fission and fusion events is controlled by dynamin family proteins; mitofusins mediate OMM fusion and Opa1 IMM fusion, whereas Drp1 in the cytosol coordinates mitochondrial fission. Fusion enables communication and sharing of matrix contents between mitochondrial compartments whilst division facilitates transport, distribution and targeted degradation of organelles (Lackner, 2014). Mitochondria are responsive to bioenergetics changes: both individual mitochondria and networks shift towards more tubular-like morphologies in a high ADP environment, as shown in Figure 1.2, although a direct mechanistic link is still lacking. Therefore, while the importance of mitochondrial networks is now accepted the precise roles and processes governing their dynamism is still unclear.

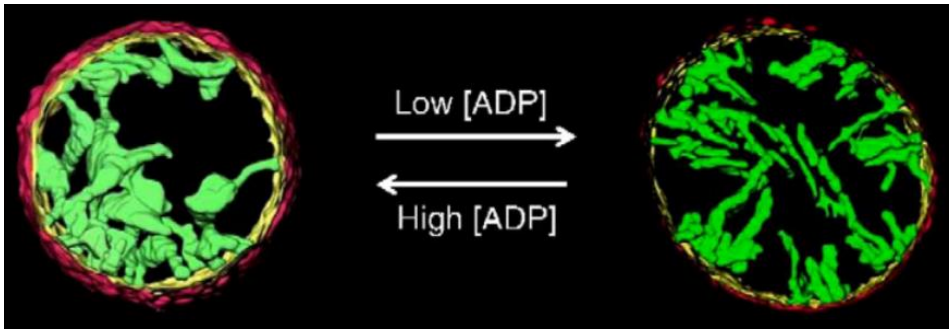


Figure 1.2 Mitochondria respond to bioenergetics changes by modification of cristae morphology. An electron tomogram is shown with OMM in red, IMM in yellow and cristae in green. High ADP levels stimulate cristae to condense to increase their internal area and form further connections with each other and the periphery of the organelle; the orthodox cristae structure persists in low ADP conditions. Figure adapted from Mannella, 2006.

Another process controlling the mitochondrial population of a cell is a specialised form of autophagy, called mitophagy. This quality control pathway is balanced with mitobiogenesis to maintain the correct levels of healthy mitochondria and closely linked to fission and fusion, as networks must be disassembled to remove damaged mitochondria. There must be recognition of impaired organelles; membrane potential depolarisation marks these mitochondria out for selective removal. Cytosolic PINK1 is known to be imported and inserted into the IMM, where it is then processed and cleaved by PARL. The cleaved product is targeted by a proteasome-dependent pathway, causing its removal from the mitochondrial membrane. However on depolarisation PINK1 is not imported or cleaved and the full-length protein accumulates in the OMM, where it recruits Parkin. Narendra *et al.* demonstrated that the physiological role of Parkin was to promote the ubiquitination of damaged mitochondria (Narendra *et al*., 2008). The interaction with PINK1 activates the ubiquitin ligase activity of Parkin, which decorates protein targets on the OMM to mark them for lysosomal degradation. The exact mechanism by which

ubiquitinated mitochondria are recognised by autophagosomes is still unclear, but it may require adaptor proteins and other E3 ligases.

1.3. Mitochondrial function

1.3.1. Energy transduction

Mitochondria play a central role in providing energy for the cell. In addition to the ETC and ATP synthase in the IMM, it is also the site of the citric acid cycle is in the mitochondrial matrix. Figure 1.3 shows how glucose is oxidised to pyruvate by glycolysis in the cytosol, which requires two ATP to be consumed in the early enzymatic stages. However the latter steps of glycolysis yield four ATP by substrate level phosphorylation and two NADH, therefore there is an overall gain of two ATP and two NADH per glucose. Although glycolysis has a low ATP yield it can be used by organisms under low oxygen conditions to maintain the ATP:ADP ratio away from equilibrium, and it is a common trait of cancer cells to rely heavily on glycolysis for their energy demands.

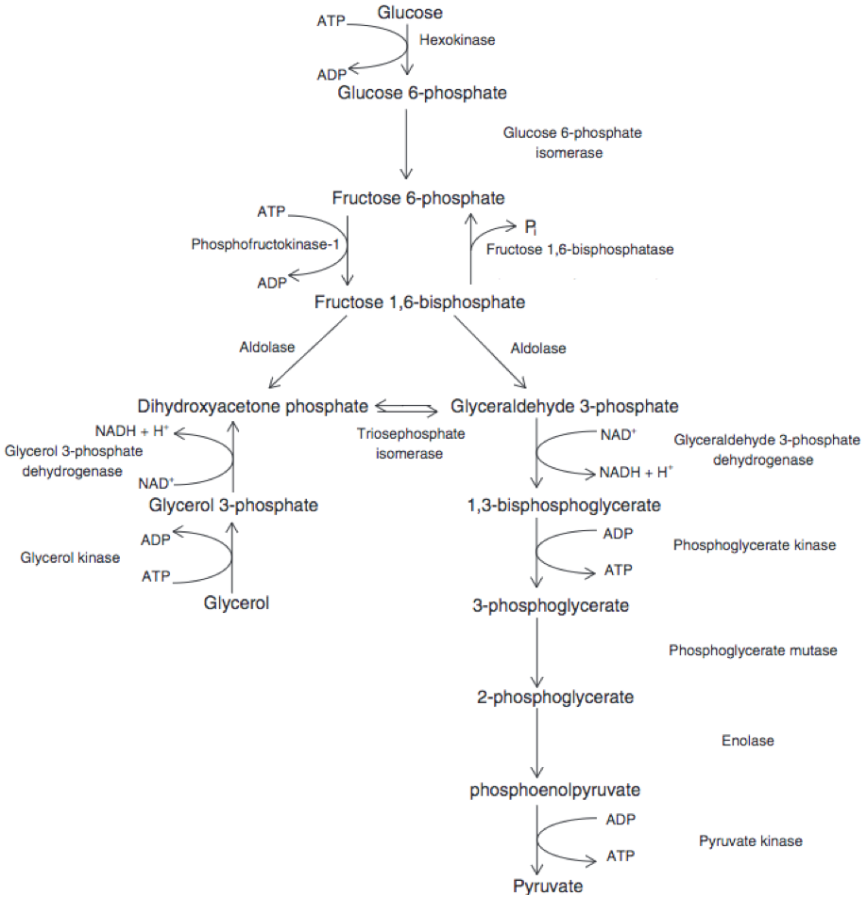


Figure 1.3 Glycolysis pathway. Steps of the glycolysis pathway and enzymes responsible for the catalysis of each step. Through isomerisation both products of aldolase formed from fructose-1,6-bisphosphate are converted to pyruvate to net yield of two ATP and two NADH for each glucose molecule. Figure adapted from Hipkiss *et al.*, 2013.

Pyruvate is a branching point for many different pathways including oxidative phosphorylation, glucose, lactate, fatty acid and amino acid synthesis. Enzymes that metabolise pyruvate are found in the mitochondrial matrix and therefore a transport-mediated process is required to move pyruvate across the IMM, although the pyruvate transporter was only identified relatively recently by two groups (Bricker *et al.*, 2012; Herzig *et al.*, 2012). For entry into the citric acid cycle pyruvate is converted into acetyl coA by the pyruvate dehydrogenase complex, additionally forming carbon dioxide and NADH. Figure 1.4 shows the citric acid cycle, comprising the activities of eight enzymes that sequentially combine acetyl coA with oxaloacetate to form citric acid and regenerate oxaloacetate. The cycle generates three NADH, two CO_2, GTP and $FADH_2$. The two decarboxylation steps (3: isocitrate to α-ketoglutatarate, and 4: α-ketoglutatarate to succinyl coA) are the irreversible exergonic stages that ensure the progression of the cycle.

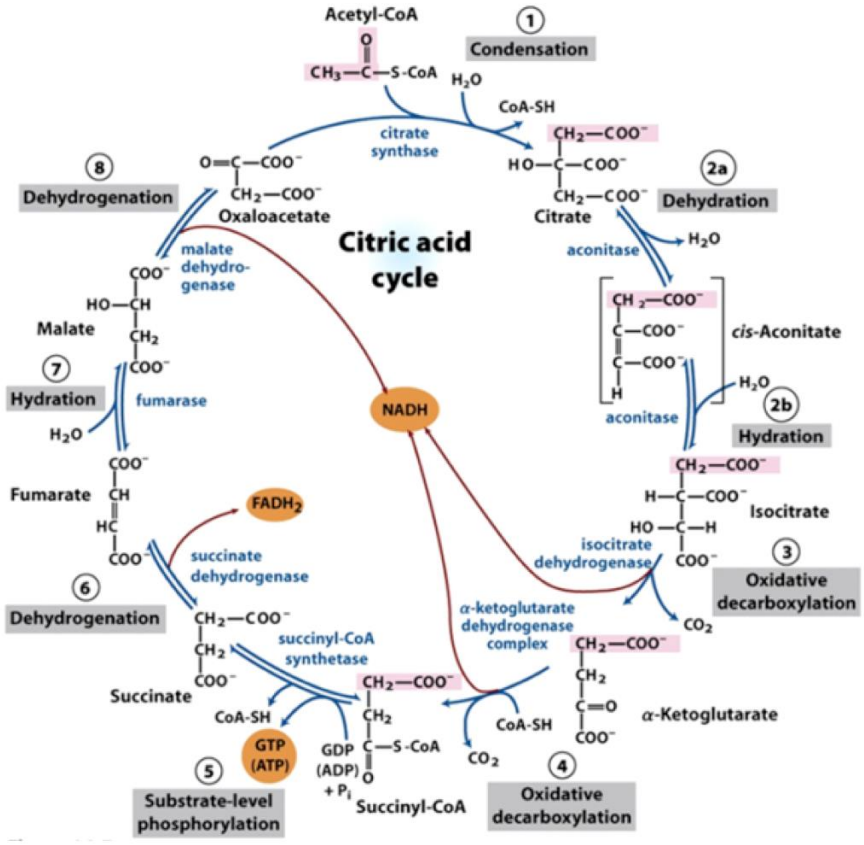


Figure 1.4 Diagram of the citric acid cycle. The enzymes are named in blue. Reducing equivalents produced are shown in orange. Figure adapted from Nelson & Cox, 2008.

Although the NADH generated by the citric acid cycle is available for use by complex I, NADH generated in the cytosol by glycolysis is not, being unable to pass across the IMM. Shuttles are utilised to move reducing equivalents across the mitochondrial membrane – the glycerophosphate shuttle and malate-aspartate shuttle. Figure 1.5A shows the glycerophosphate shuttle, which involves the conversion of the glycolytic intermediate dihydroxyacetone phosphate into glycerol-3-phosphate by glycerol-3-phosphate dehydrogenase, using NADH. The reverse reaction is carried out by another

glycerol-3-phosphate dehydrogenase, a flavoprotein localised to the IMM. This results in the reduction of FAD to $FADH_2$, which can then reduce ubiquinone to ubiquinol for entry of electrons into the electron transport chain. The shuttle bypasses complex I and therefore results in fewer protons pumped and less ATP generated, since this process is irreversible it can be utilised under all cellular conditions. The other shuttle uses oxaloacetate in the cytoplasm, which is converted into malate by malate dehydrogenase, coupled to NADH oxidation. Figure 1.5B shows that malate enters into mitochondria through an exchanger, which exports α-ketoglutatarate. Malate is then oxidised back to oxaloacetate in the matrix by a mitochondrial malate dehydrogenase, forming NADH that then enters the ETC at complex I, maximising ATP formation. Oxaloacetate is returned to the cytoplasm through a separate transporter, which exchanges aspartate for glutamate (α-ketoglutatarate and oxaloacetate are the α-keto acid equivalents of glutamate and aspartate). This is necessary to allow the net movement of electrons from one side of the membrane to the other. Overall both shuttles result in the transport of electrons into the mitochondria, without a change in metabolites.

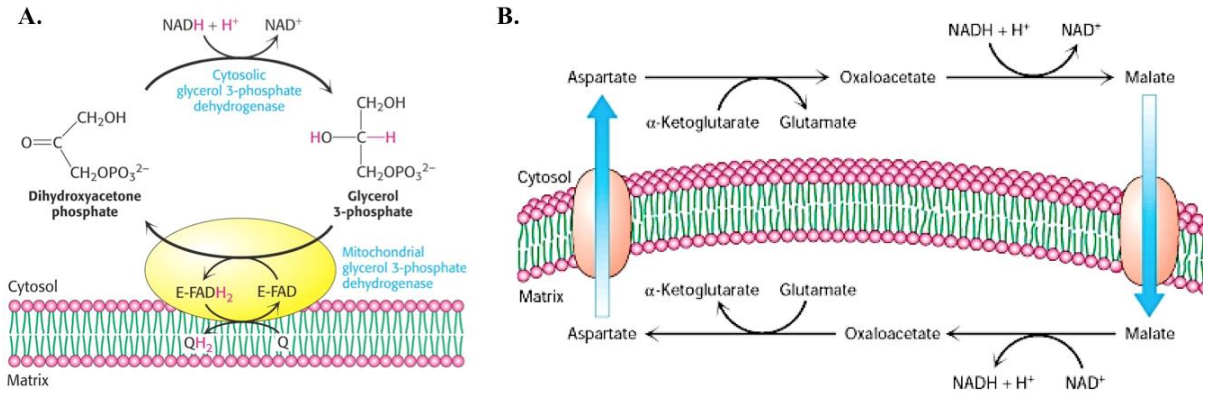


Figure 1.5 The mitochondrial glycerophosphate and malate-aspartate shuttles. A. The glycerophosphate shuttle. Dihydroxyacetone phosphate is converted to glycerol-3-phosphate by glycerol-3-phosphate dehydrogenase in the cytosol, consuming NADH. The reverse reaction can then be carried out by another glycerol-3-phosphate dehydrogenase, a flavoprotein localised to the IMM (yellow). FAD is reduced to $FADH_2$, which then reduces ubiquinone to ubiquinol within the IMM. Electrons can then enter the electron transport chain. **B.** The malate-aspartate shuttle. Oxaloacetate in the cytoplasm is converted into malate by malate dehydrogenase, requiring NADH. Malate enters the mitochondria through an exchanger (orange). Malate is oxidised back to oxaloacetate in the matrix by a mitochondrial malate dehydrogenase, forming NADH that may then enter the electron transport system at complex I, maximising ATP formation. Oxaloacetate is returned to the cytoplasm through a separate transporter via aspartate. Reducing equivalents can therefore be transferred across the IMM without net movement of metabolites. Figures adapted from Berg *et al*., 2002.

NADH, as a source of electrons for complex I, is also formed by other mitochondrial pathways. Beta-oxidation of fatty acids involves cyclic rounds of two-carbon removal and yields one NADH, one $FADH_2$ and one acetyl coA with each round. The NADH produced is available for re-oxidation by complex I whereas the $FADH_2$ is the cofactor of the flavoprotein ETF (Gregersen, 1985). The reduced ETF in turn reduces electron transfer flavoprotein dehydrogenase (ETFDH), which then reduces ubiquinone, allowing these electrons to enter the respiratory chain through the quinone pool. Another

key source of electrons is the oxidation of ketone bodies, where the first step is the NADH-dependent oxidation of β-hydroxybutyrate (Cotter *et al*., 2013). Additionally the end product is acetyl coA, so in the same way as β-oxidation of fatty acids, NADH is provided to the electron transport chain for complex I and acetyl coA sequestered by the citric acid cycle, all contributing towards ATP production. Energy production through ketone body derived reducing equivalents is utilised by some tissues under stress conditions, particularly the brain under low glucose; it is unable to utilise fatty acids, which do not pass the blood-brain barrier. The heart also derives NADH from ketone oxidation whereas the liver exports ketone bodies, being unable to oxidise them due to lack of succinyl-CoA: 3-ketoacid transferase (Orii *et al.*, 2008).

The requirement of electrons for oxidative phosphorylation and their multiple sources means that the ETC is intricately linked to the rest of mitochondrial metabolism and there is a reciprocal requirement from each to sustain turnover. Complex I cannot function without a NADH supply and if its activity is impaired then re-oxidation of NADH is limited for use by the citric acid cycle, fatty acid oxidation and ketone body oxidation. In this way complex I serves as an important connection between many mitochondrial pathways.

Mitochondria also perform many other functions within the cell, some of which are summarised below.

1.3.2. Calcium homeostasis

Mitochondria are an important Ca^{2+} sink that can store and release the ion in response to changes in the cytoplasm. Ca^{2+} are imported across the OMM by a large voltage gated channel, VDAC, and across the IMM through two pathways – either using a uniporter moving the ions down their electrical gradient (established by ETC proton pumping) or through exchangers. The Na^{+}/Ca^{2+} and the H^{+}/Ca^{2+} exchangers are electroneutral antiporters and prevent electrochemical equilibrium. The molecular identity of the uniporter (MCU) was only established in 2011 (Baughman *et al*., 2011). Mitochondria, along with the endoplasmic reticulum, enable management of intracellular $[Ca^{2+}]$, which provides a signalling mechanism, allowing mitochondria to respond to the energy requirements of the cell. Three mitochondrial dehydrogenases, pyruvate dehydrogenase, isocitrate dehydrogenase and α-ketoglutatarate dehydrogenase, are regulated by calcium. Increased work causes cytosolic $[Ca^{2+}]$ to rise and the matrix $[Ca^{2+}]$ also increases, leading to the stimulation of these mitochondrial dehydrogenases to increase NADH production, enhancing electron transport through the respiratory chain and ultimately increasing ATP production. Additionally, it has been proposed that ATP synthase can also be activated by calcium (Territo *et al*., 2000).

1.3.3. Iron sulphur synthesis

The biogenesis of iron sulphur cofactors is also partially undertaken in mitochondria. [2Fe-2S], [3Fe-4S] or [4Fe-4S] clusters and are required by proteins for a range of functions including electron transfer catalysis, and they are important sulphur donors in lipoic acid and biotin synthesis (Stehling & Lill, 2013). Their synthesis and incorporation into apoproteins involves around 30 proteins in the mitochondria and cytosol. Biosynthesis starts in the matrix with the iron-sulphur assembly machinery, comprising more than 15 components, and the cofactor may then be retained in the mitochondria or exported for maturation of non-mitochondrial Fe-S proteins. Impairment of FeS protein assembly is associated with severe pathology including neurodegenerative, metabolic and haematological conditions, which often display accumulation of iron in the mitochondria (Stehling & Lill, 2013).

1.3.4. Apoptosis

Mitochondria also play a central role in programmed cell death, which may be in response to external or internal factors. This process is tightly regulated by Bcl-2 family proteins, with both anti-apoptotic and pro-apoptotic members (Wang & Youle, 2009). The balance of these factors controls the permeability of the OMM and release of caspases and cytochrome c into the cytosol. Additionally there is sustained IMM permeability due to the opening of the still elusive permeability transition pore (Izzo *et al*., 2016). Therefore mitochondria contain the signals necessary to bring about irreversible cell disassembly: and so the retention of apoptosis regulating molecules in mitochondria is essential, along with the fidelity of the signalling pathway that brings about their release.

1.4. The electron transport chain

The respiratory chain coordinates electron transfer from NADH to oxygen and couples it to proton translocation to maintain Δp. It is composed of four protein complexes. Complex I is the largest complex, it obtains electrons from oxidation of NADH in the mitochondrial matrix, and passes them to its other substrate, ubiquinone (Q), a hydrophobic acceptor present in the IMM. The detailed structure and function of complex I will be discussed separately in section 1.5.

1.4.1. Complex II

Complex II (succinate dehydrogenase) catalyses the oxidation of succinate to fumarate and reduces ubiquinone to ubiquinol by a two-electron transfer. The enzyme is both an integral part of the respiratory chain and participates in the citric acid cycle, but unlike complex I does not pump protons. Complex II is formed of four subunits, which are all nuclear-encoded and therefore is the only ETC complex without any mitochondrially-encoded subunits. SDHA and SDHB are the hydrophilic core components and

contain the redox cofactors shown in Figure 1.6. FAD is covalently bound to SDHA, and the three iron-sulphur clusters that facilitate electron transfer to ubiquinone are coordinated by SDHB. These are a 2Fe-2S cluster, which sits ~12 Å from the FAD, followed by a 4Fe-4S and a 3Fe-4S cluster (Sun *et al*., 2005). Additionally, complex II contains a haem within the membrane domain; but its exact role is unclear, although it is not thought to be involved in catalysis. A mutant in *Saccharomyces cerevisiae* that lacks the haem was not catalytically comprised (Oyedotun *et al*., 2007). Due to the small difference in midpoint potential between succinate and ubiquinone (~0.03 V) there is little energy released upon oxidation of succinate and therefore complex II is unable to pump protons.

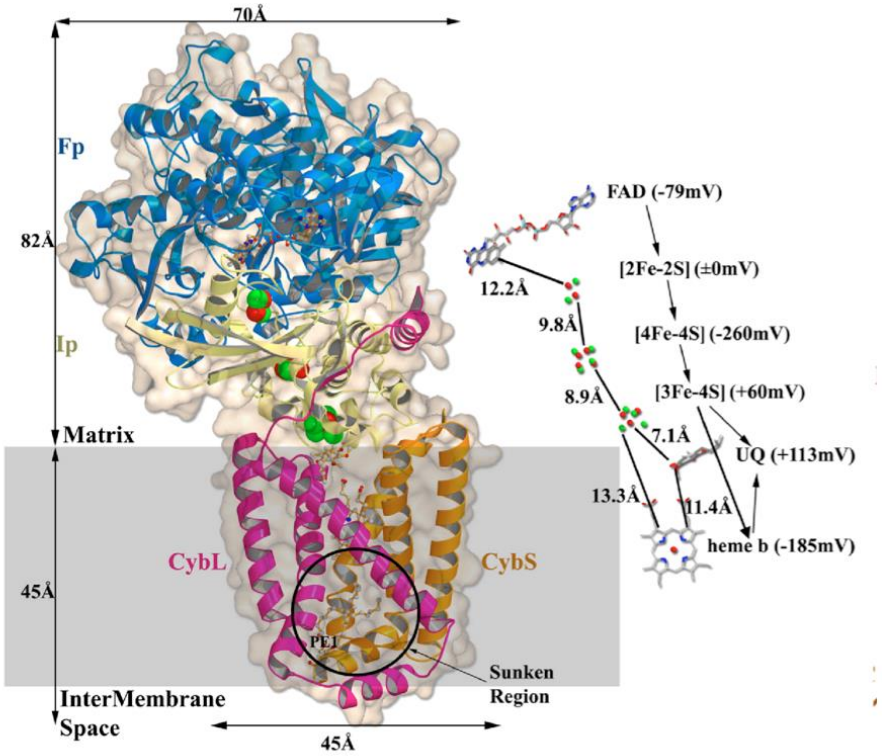


Figure 1.6 Structure of succinate dehydrogenase from *Sus scrofa* at 2.4 Å resolution. The crystal structure of complex II is shown embedded in the IMM. SDHA is shown in blue ribbons, SDHB in cream, with transmembrane SDHC and SDHD in pink and orange. The cofactors, their midpoint potentials and distances within the structure are shown to the right of the structure. Figure adapted from Sun *et al*., 2005.

The reduction of ubiquinone requires the uptake of two protons to form ubiquinol (QH_2), which is confined to the IMM due to its hydrophobicity. QH_2 acts as a carrier, shuttling electrons from complexes I and II to complex III. The redox state of the ubiquinone pool is also affected by the activities of other proteins, including the electron transfer flavoproteins (ETFs), such as ETF dehydrogenase in the β-oxidation pathway for fatty acids.

1.4.2. Complex III

Complex III (cytochrome bc_1 complex) is the next component of the ETC: it contributes to Δp, but via a modified Q-cycle rather than through direct proton pumping (Crofts *et al.*, 1983). The complex has 11 subunits, although only three are necessary for catalysis and conserved in all systems; the others may aid assembly and stability. The redox cofactors are: a 2Fe-2S Rieske centre, a c-type haem and two b-type haems. The two b-haems have differing redox potentials, which is central to the mechanism of the enzyme – b_L with an $E_{m,7}$ ~ -30mV is located adjacent to the intermembrane space, whereas b_H with $E_{m,7}$ ~ +100mV is proximal to the matrix (Xia *et al.*, 2013). They connect the two binding sites for ubiquinone: Q_P is found close to the positive (intermembrane) face and Q_N near to the negative (matrix) side.

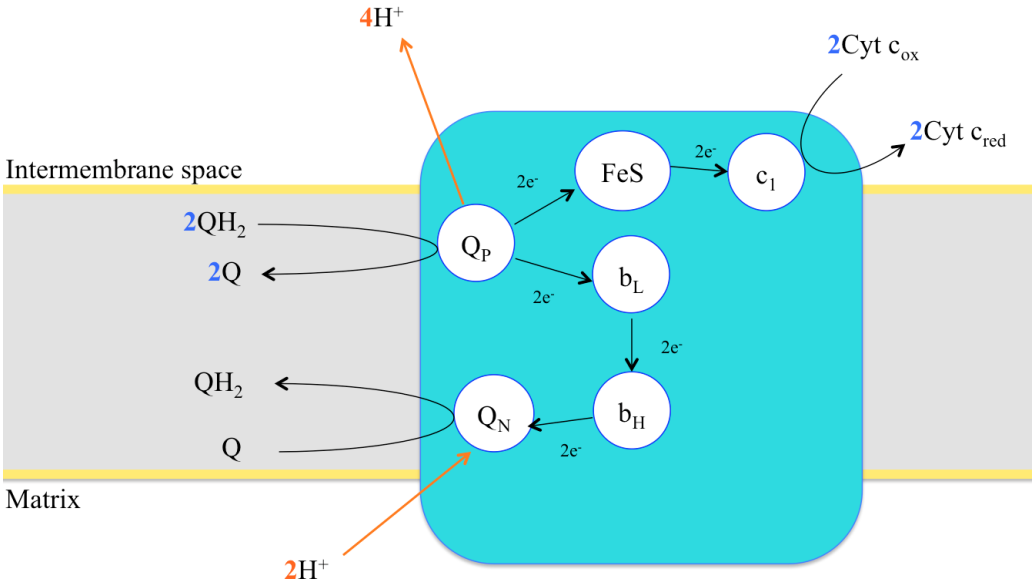


Figure 1.7 Mechanism of ubiquinone oxidation by complex III. The diagram shows the net result of two proton motive Q-cycles by the complex III monomer. The FeS cluster is located on the Rieske protein, the c-type haem (c_1) on the cytochrome c subunit ($cytc_1$) and the b-type haems (b_L and b_H) on the cytochrome b subunit. Arrows show electron movement along the two pathways – high potential through the Rieske protein to the cytochrome c subunit and onto cytochrome c and low potential pathway along the b-type haems to the Q_N site to reduce ubiquinone.

Figure 1.7 shows the mechanism of ubiquinone oxidation – first, a molecule of QH_2 binds at the Q_P site, which neighbours the Rieske protein, such that one of the QH_2 hydroxyls forms a hydrogen bond to a histidine coordinating the Fe-S centre. Oxidation of QH_2 then occurs in two stages. The first electron is transferred to the Rieske iron-sulphur cluster and two protons are released into the intermembrane space, forming a free radical semiquinone ($Q^{\bullet -}$) at the Q_P site (Trumpower, 1990). The second electron is then transferred to the b_L haem, also near the positive side of the membrane. The semiquinone is unstable so the second oxidation is energetically favourable. Overall, the bifurcation reaction sends one electron to the high potential chain and one to the low potential chain. However the semiquinone may also reduce oxygen in an unwanted side reaction (Trumpower, 1990). The electron received by the Rieske protein

passes down the high potential chain to $cytc_1$ and on to cytochrome c (top pathway in Figure 1.7). The low potential pathway is from the $Q^{\bullet-}$ to the b_L haem and b_H haem (lower pathway). A $\Delta\Psi$ of +150 mV opposes the electron transfer from b_L to b_H, from the positive to negative side of the membrane, but the drop in reduction potential balances this energetically unfavourable reaction, allowing the electron to retain its original energy (Nicholls & Ferguson, 2013). At the second quinone-binding site Q_N, close to b_H at the matrix side of the membrane, Q binds and accepts the electron from b_H, forming $Q^{\bullet-}$. The $Q^{\bullet-}$ is stable in the Q_N, such that the first and second reduction event to occur with similar energy. To complete the cycle another QH_2 is oxidised at Q_P with one electron passing to cytochrome c through the high potential pathway and the other passing from b_L to b_H along the low potential pathway. The second electron reduces $Q^{\bullet-}$ to QH_2 and the two protons required are taken up from the matrix, allowing QH_2 to return to the pool. Overall one ubiquinol is reduced by two cytochrome c molecules, coupled to the translocation of four protons.

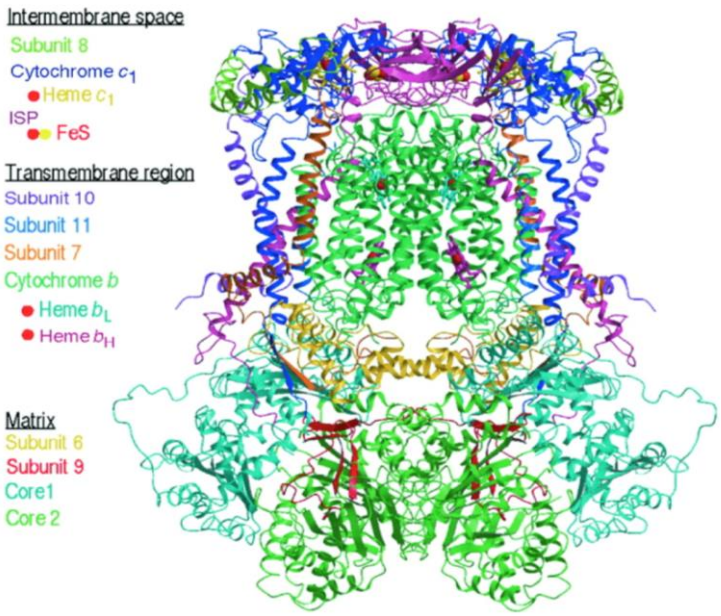


Figure 1.8 Structure of cytochrome bc_1 complex homodimer from *Bos taurus* at 3.0 Å resolution. The crystal structure of complex III shows its native state as a homodimer, with subunits highlighted by colour. Figure adapted from Iwata *et al.*, 1998.

The structure of complex III was important for confirming and increasing understanding of the mechanism, particularly relating to the bifurcation at the Q_P site. The different pathways for each electron have been further elucidated by the observation that the globular head of the Rieske protein – containing the 2Fe-2S cluster – changes its position. When the cluster is reduced the head docks onto $cytc_1$ to allow transfer of the electron, after which the affinity for the cytochrome c_1 subunit is decreased and the head returns to the Q_P site, where it is electrostatically stabilised. The movement occurs across a 20 Å range on a sub-millisecond time-scale, which matches electron transfer rates (Yu, *et al.*, 2008). Additionally, as the distance limit for rapid electron transfer is 14 Å, when the Rieske head is bound to

cytc$_1$ it is too far away to accept the second electron, which is directed down the low potential chain. As shown in Figure 1.8, complex III acts as a dimer with cooperation between the monomers: the structure shows there is a crossover, with the head group of the Rieske protein on one subunit interacting with Q_P and cytc$_1$ in the other (Iwata *et al*., 1998).

Cytochrome c then acts as an electron shuttle; it diffuses through the intermembrane space carrying a single electron on its c-type haem, which it transfers to complex IV in the final step of the ETC.

1.4.3. Complex IV

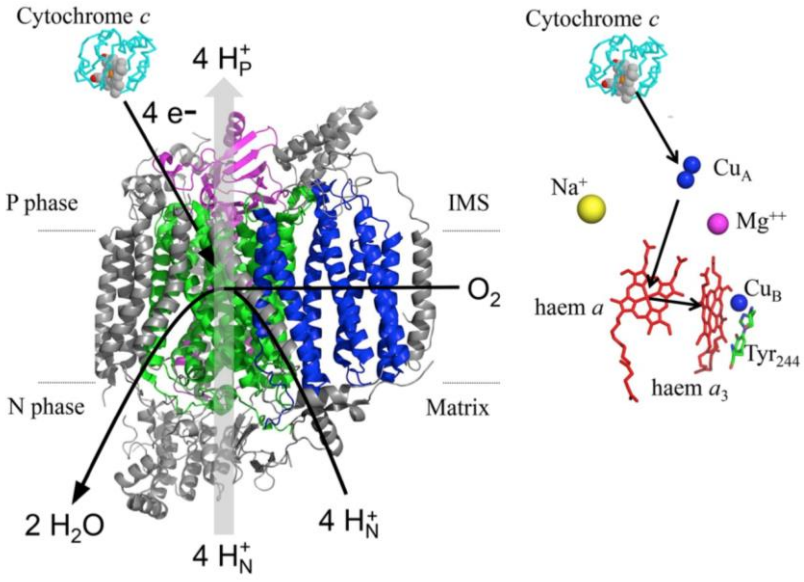


Figure 1.9 Structure of cytochrome c oxidase from *Bos taurus* at 1.8 Å resolution. The crystal structure of complex IV is shown with core subunit I in green, II purple and III in blue. The path of electron transfer is shown on the right of the structure. Figure adapted from Rich, 2017.

Figure 1.9 shows the structure of mammalian complex IV, which comprises 13 subunits. Three are mitochondrially encoded (COX I, II, III). Catalysis requires four cytochrome c to sequentially donate four electrons to reduce oxygen to two molecules of water. Cytochrome c in the intermembrane space transfers an electron to the dinuclear Cu_A centre of subunit II, it then moves to the low spin haem a and onto the haem a3/Cu_B binuclear centre (BNC). Oxygen enters the BNC at subunit III, where it is reduced to two water molecules with the uptake of four protons from the matrix. Upon each electron reduction of oxygen one proton is moved across the membrane into the IMS and a second substrate proton is taken up by the oxygen. Although the protons and electrons utilised for the reduction of oxygen do not completely cross the membrane they do contribute to $\Delta\Psi$ as they move towards the positive and negative sides of the membrane respectively and cross part of the dielectric distance (Belevich *et al*., 2007). The energy released upon reduction of oxygen drives the pumping of four protons into the IMS.

One of the least-understood aspects of complex IV biochemistry is the movement of protons, both as substrates to reduce oxygen and those pumped into the IMS. Two proton channels have been designated the K and D channels, named for particular lysine or aspartate residues. The K-channel connects the matrix to the catalytic tyrosine of the BNC and the D-channel leads from an aspartic aid at the matrix side to a region equidistant to the haems (Yoshikawa *et al*., 2011). Additionally a H-channel was first described in bovine cytochrome c oxidase and is more complex (Muramoto *et al*., 2010). It has been difficult to ascertain whether specific channels are related to the movement of substrate and translocated protons or whether there is some cross over, although it has been suggested that the K-channel feeds two substrate protons to the active site, whereas the D-channel transfers both types (Rich, 2017).

1.4.4. Complex V

The Δp generated by complexes I, III and IV is then harnessed to generate ATP. Mammalian ATP synthase (complex V) is a 580 kDa complex with 15 different subunits. It comprises a hydrophobic F_o domain in the membrane and a hydrophilic F_1 domain that extends into the matrix, as depicted in Figure 1.10. The F_1 domain region is a hetero-hexamer of α and β subunits, which may bind ATP, ADP and P_i. The conformational changes at the catalytic site are controlled by the γ subunit, which rotates non-symmetrically in the centre to bring about the three-site alternating binding site mechanism. The peripheral stalk anchors the F_1 domain to the membrane embedded F_o domain, made up of the b subunit, F6, d and oligomycin-sensitivity conferring (OSCP) subunits, to prevent rotation of the $\alpha_3\beta_3$ assembly. Recently the intact F-ATPase from *Paracoccus denitrificans* has been crystallised, shown in Figure 1.11, providing additional insight into the coupling of the proton motive force to ATP synthesis (Morales-Rios *et al*., 2015).

The F_o domain comprises a ring of c subunits, with the copy number depending on the species; the human and bovine enzymes have c-8 rings (Watt *et al*., 2010). Each subunit has two helices separated by a loop and an ionisable residue. The a subunit is thought to provide a proton entrance at the IMS side and an exit channel at the matrix side, through which protons access an ionisable residue of the c subunit. Once neutralised the residue moves the subunit away from the channel and rotates the c-ring assembly, the proton is then lost through the exit half channel of subunit a (Nicholls & Ferguson, 2013). One complete rotation of the c-ring cause the full rotation of the $\alpha_3\beta_3$ assembly and generates three ATP. The inhibitor protein IF1 binds in a pH dependent manner and prevents ATP synthase operating in reverse. At lowered matrix pH it binds in response to a decreased Δp to prevent ATP hydrolysis (Gledhill *et al*., 2007).

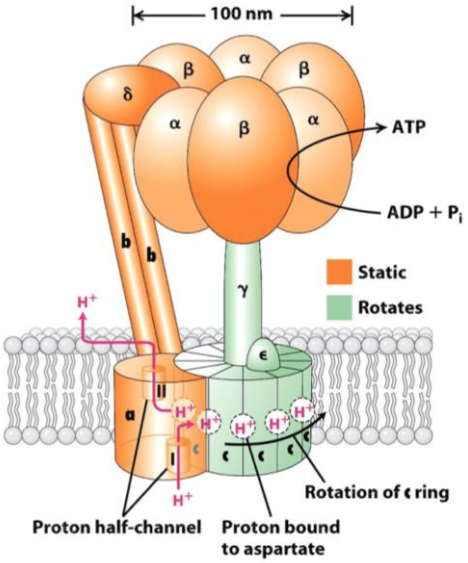


Figure 1.10 Schematic representation of *E. coli* ATP synthase. Subunits are labelled and shows the hydrophilic F_1 domain extending into the matrix. Orange subunits are static and green subunits rotate during ATP synthesis. The principle is the same for the mammalian enzyme. Figure adapted from Lodish *et al*., 2008.

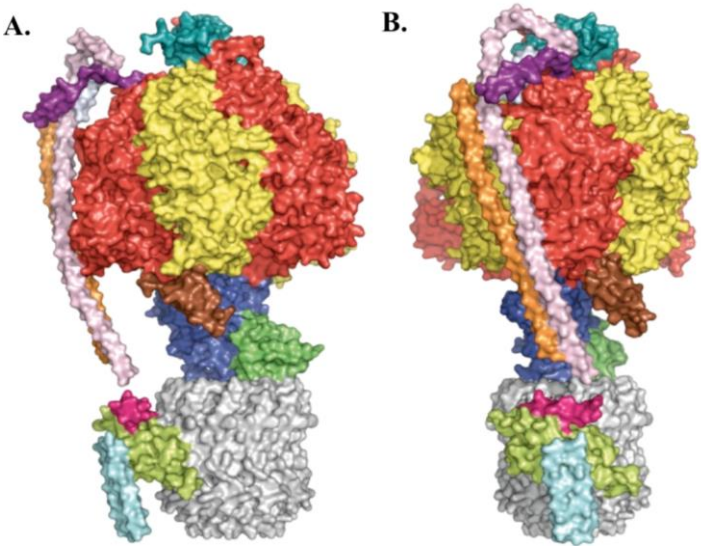


Figure 1.11 Crystal structure of ATP synthase from *P. denitrificans* at 4.0 Å resolution, with inhibitor protein bound. A. Side view of the enzyme with **B.** rotated 90 ° to **A.** The most complete structure of ATP synthase showing the new features of the peripheral stalk and same unique fold in the bacterial a-subunit as in the mitochondrial ATP6 subunit. Helix-1 of the inhibitor is shown in brown. Figure adapted from Morales-Rios *et al*., 2015.

1.5. Complex I

1.5.1. Subunits and structure

Mammalian complex I (NADH:ubiquinone oxidoreductase) is the largest of the ETC enzymes, with 45 subunits contributing to a mass of ~1 MDa. The core enzyme comprises 14 subunits, which are retained across all species containing complex I. They are segregated into the membrane arm, with seven transmembrane subunits all encoded by mtDNA (ND subunits) and seven hydrophilic subunits encoded in the nucleus and imported to the mitochondria giving the distinctive L-shape (Hirst, 2013). Table 1.1 shows the subunit nomenclature for different species, demonstrating that whilst mammalian complex I contains 45 subunits, the yeast (*Yarrowia lipolytica*) enzyme shares only some of the supernumerary subunits found in mammals – but also has four species specific subunits. Bacterial and archaeal (*Escherichia coli* and *Thermus thermophilus*) complex I represent the core enzyme only.

Domain	*Bos tarus*	*Homo sapiens*	*Yarrowia lipolytica*	*Thermus thermophilus*	*Escherichia coli*
Hydrophobic arm	ND1	ND1	NU1M	Nqo8	NuoH
	ND2	ND2	NU2M	Nqo14	NuoN
	ND3	ND3	NU3M	Nqo7	NuoA
	ND4	ND4	NU4M	Nqo13	NuoM
	ND4L	ND4L	NULM	Nqo11	NuoK
	ND5	ND5	NU5M	Nqo12	NuoL
	ND6	ND6	NU6M	Nqo10	NuoJ
Hydrophilic arm	24 kDa	NDUFV2	NUHM	Nqo2	NuoE
	49 kDa	NDUFS2	NUCM	Nqo4	NuoCD
	30 kDa	NDUFS3	NUGM	Nqo5	
	51 kDa	NDUFV1	NUBM	Nqo1	NuoF
	75 kDa	NDUFS1	NUAM	Nqo3	NuoG
	PSST	NDUFS7	NUKM	Nqo9	NuoB
	TYKY	NDUFS8	NUIM	Nqo6	NuoI
Supernumerary subunits	10 kDa	NDUFV3			
	13 kDa	NDUFS6	NUMM		
	15 kDa	NDUFS5	NIPM		
	18 kDa	NDUFS4	NUYM		
	39 kDa	NDUFA9	NUEM		
	42 kDa	NDUFA10			
	ACP	NDUFAB1α	ACPM		
	ACP	NDUFAB1β			
	AGGG	NDUFB2			
	ASHI	NDUFB8	NIAM		
	B8	NDUFA2	NI8M		
	B9	NDUFA3	NI9M		
	B12	NDUFB3	NB2M		
	B13	NDUFA5	NUFM		
	B14	NDUFA6	NB4M		
	B14.5a	NDUFA7	NUZM		
	B14.5b	NDUFC2			
	B14.7	NDUFA11	NUJM		
	B15	NDUFB4	NB5M		
	B16.6	NDUFA13	NB6M		
	B17	NDUFB6			
	B17.2	NDUFA12	N7BM		
	B18	NDUFB7	NB8M		
	B22	NDUFB9	NI2M		
	ESSS	NDUFB11	NESM		
	KFYI	NDUFC1			
	MNLL	NDUFB1			
	MWFE	NDUFA1	NIMM		
	PGIV	NDUFA8	NUPM		
	PDSW	NDUFB10	NIDM		
	SGDH	NDUFB5			
Species specific subunits			NEBM		
			NUNM		
			NUUM		
			NUXM		

Table 1.1 Subunit composition of complex I and nomenclature used across different species. The names given to the subunits of complex I in species commonly used for its study are shown. Only core subunits are present in complex I from *T. thermophilus* and *E. coli*, whereas *Y. lipolytica* contains 20 of the supernumerary subunits, of which there are 32 in total, shared by bovine, human and other mammalian complex I (Berrisford *et al*., 2016; Angerer *et al*., 2011).

The first structures of complex I were from bacterial species, so they only provided information on the core subunits. *E. coli* and *T. thermophilus* have both been used as model systems. The hydrophilic domain of *T. thermophilus* complex I was resolved to 3.3 Å by Sazanov and co-workers in 2006, closely followed by the membrane domain structure of *E. coli* at 3.0 Å and later a high resolution structure of the complete *T. thermophilus* enzyme at 3.3 Å in 2013, Figure 1.12 (Sazanov & Hinchliffe, 2006; Efremov & Sazanov, 2011; Baradaran *et al.*, 2013). These structures elucidated the FMN active site; the arrangement of iron sulphur clusters and indicated a putative Q-binding site. Additionally, an unusually positioned 2Fe-2S cluster (2Fe[24]) was observed on the opposite side of the flavin. The membrane domain from *E. coli* also highlighted a transverse helix, running along the top of the membrane domain. Figure 1.12 shows the high resolution *T. thermophilus* complex, from which the putative Q-binding site was identified as a long tunnel, with a narrow entry point for the quinone. A possible structural basis for coupled proton pumping was also indicated. Three antiporter-like subunits - Nqo12, Nqo13, Nqo14 (ND5, ND4 and ND2) - share homology with Mrp antiporters and pump protons, but a mechanism is required to couple changes at the Q-site to proton pumping by complex I. The *T. thermophilus* structure highlighted that subunit Nqo8 (ND1), which sits in the membrane domain at the base of the hydrophilic arm, contains a similar fold to the antiporter-like subunits, despite a low sequence similarity. It contains a mixture of long and short helices tilted relative to the membrane. This region can be superimposed onto the half-channels of the antiporter-like subunits and the second half-channel was proposed to be formed by Nqo10 and Nqo11 (ND6 and ND4L) (Baradaran *et al.*, 2013). Charged residues were identified in Nqo8 that mirrored those found in the antiporter-like subunits, along with an additional network creating a 'funnel' to connect the charged midline residues to the Q-site, which may facilitate transmission of conformational changes. Therefore a fourth proton channel was elucidated from the *T. thermophilus* structure, named the E-channel for the abundance of glutamate residues, and is formed by Nqo8, Nqo10 and Nqo11 (ND1, ND6 and ND4L). The channel is made up of many conserved residues, which are crucial for complex I activity. Consequently the presence of four putatitve proton channels in the *T. thermophilus* structure supported the proposed complex I proton pumping stoichiometry of four, rather than three (Baradaran *et al.*, 2013).

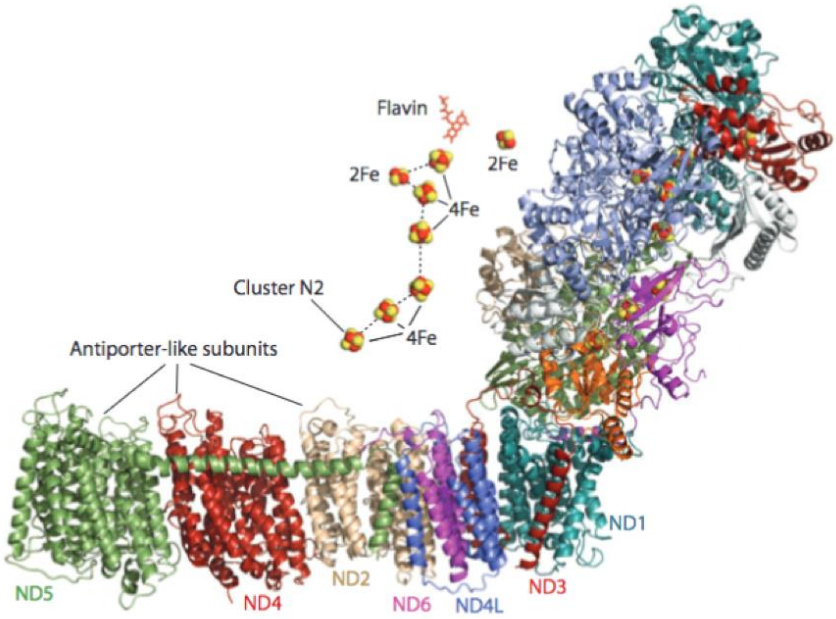


Figure 1.12 Structure of complex I from *Thermus thermophilus* at 3.3 Å resolution. The first complete structure of complex I showing all core subunits, shared with mammalian complex I (Baradaran *et al.*, 2013: PDB code 4HEA). Mitochondrially encoded core subunits are coloured to match their labels and the chain of redox cofactors is shown next to the hydrophilic domain. The 2Fe-2S iron sulphur cluster to the right of the bound flavin (as shown) is isolated from the rest of the chain and not thought to be involved in electron transfer. Figure adapted from Hirst, 2013.

Structures of the mammalian enzyme initially proved more elusive, with crystallography impeded by the large size of complex I and its membrane protein characteristics. However its size, along with its asymmetry, enabled single particle cryo-electron microscopy (cryo-EM) to be utilised and in 2014 Hirst and co-workers determined the bovine enzyme structure to 5 Å, allowing the 14 core subunits and a further 14 supernumerary subunits to be assigned (Vinothkumar *et al*., 2014). Figure 1.13 shows the bovine complex I structure. The conserved bovine subunits showed high similarity to the bacterial core components and Figure 1.13B highlights that the supernumerary subunits form a cage-like covering over the surface of the enzyme, possibly to stabilise the structure, whilst leaving the NADH binding site accessible to substrate. With further refinement more subunits were fitted to the structure and in 2016 all 45 subunits were located and modelled in a 4.2 Å structure (Zhu *et al*., 2016). This structure not only strengthened the identity and properties of the Q-binding channel and possible proton pumping coupling mechanism but also demonstrated distinct conformational states of complex I, designated the deactive and active forms, with the elucidation of changes in the long matrix loop of ND3.

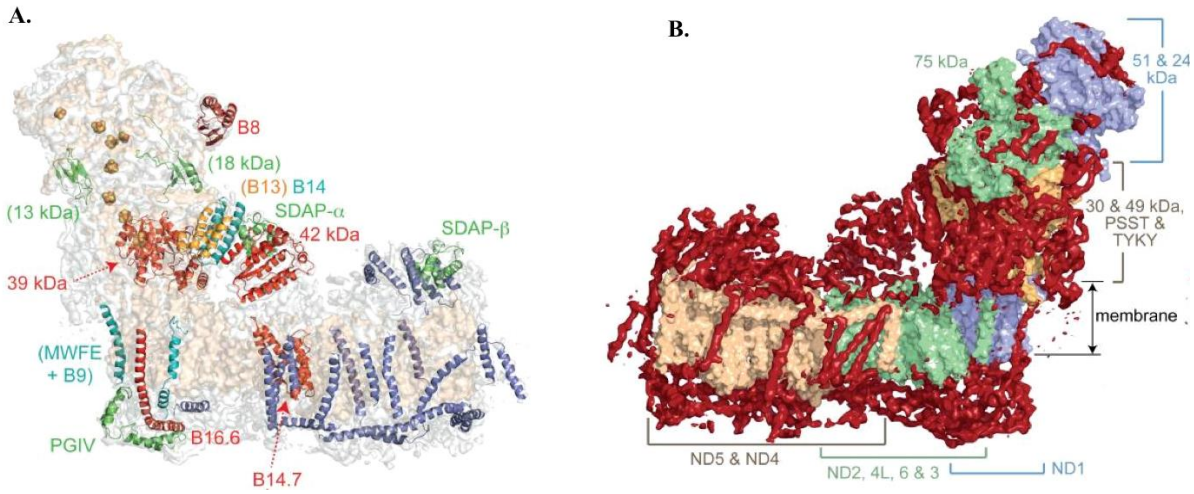


Figure 1.13 Structure of complex I from *Bos taurus* at 5.0 Å resolution. A. Assignment of supernumerary subunits coloured to match labels, those subunits on the reverse side of the structure are labelled with dashed arrows. The electron density map is shown in grey and the surface model of core subunits in wheat. The remaining supernumerary subunits were assigned to the refined structure in 2016 (Zhu *et al*., 2016). B. Location of the supernumerary subunits around the core enzyme. The 14 core subunits are shown as a surface representation in colour, enclosed by the density of the supernumerary subunits in red. These form a supportive case around the core protein but are missing from regions required for substrate access, such as the 51 kDa subunit for NADH binding. Figure adapted from Vinothkumar *et al.*, 2014.

Further structures using cryo-EM have added to increasingly detailed knowledge of the structure, including complex I from ovine-heart mitochondria (Fiedorczuk *et al*., 2016). The different states of the enzyme have been further explored - a 4.1 Å resolution structure of the deactive bovine enzyme showed that in the absence of substrate the enzyme adopts a resting state, with a disordered Q-binding site, which requires the reintroduction of substrate for reactivation (Blaza *et al.*, 2018). Most recently complex I from mouse heart mitochondria has been resolved to 3.3 Å in the active state, Figure 1.14. Comparison with the deactive structure revealed some conformational changes, especially in a potentially important π-bulge in subunit ND6. Differences have also been noted from previous bacterial structures, altered helical geometry in ND4 may have consequences for proton translocation due to the positioning of charged residues (Agip *et al*., 2018). Along with different functional states, complex I is also found in supercomplexes, forming strong contacts with complex III. Respirasomes are comprised of complexes I, III and IV with ratios $I_1III_2IV_{1-4}$ (Lenaz *et al.*, 2016). High-resolution structures of supercomplexes have been obtained, including from porcine-heart mitochondria and cultured human cells, which have identified the key subunit contacts between complexes: notably NDUFB9 and NDUFA11 interact with UQCR1 and UQCRQ of complex III, respectively (Wu *et al*., 2016; Guo *et al*., 2017). There is much speculation on the function of supercomplexes, including aiding stability of complex I, allowing substrate channeling between complexes I and III and controlling ROS production; however, this area of ETC function remains contentious and still requires further biochemical assessment to determine the advantages of these multi-complex arrangements (Sousa *et al*., 2018).

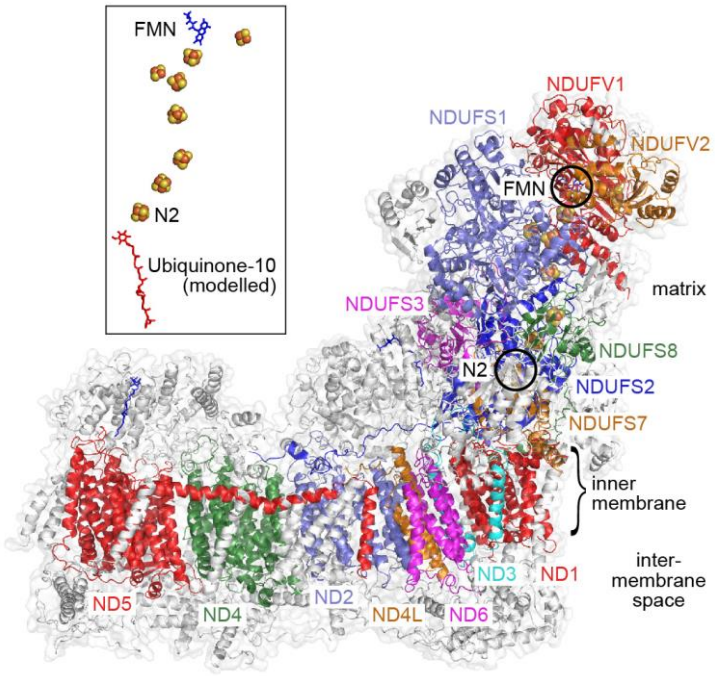


Figure 1.14 Structure of complex I from *Mus musculus* at 3.3 Å resolution. An overview of the best-resolved structure of complex I to date, with the 14 core subunits in colour and supernumerary subunits in white. The site of FMN binding and position of the terminal iron sulphur cluster N2 are highlighted. The redox cofactor chain is shown inset above. Figure adapted from Agip *et al*., 2018.

1.5.2. The catalytic mechanism of complex I

1.5.2.1. NADH oxidation at the flavin-binding site

The first step of ubiquinone reduction by complex I is the oxidation of NADH by electron transfer to bound FMN on NDUFV1. NADH binds with its nicotinamide ring positioned over the isoalloxazine moiety of FMN, as occurs in many oxidoreductases (Berrisford & Sazanov, 2009; Ferguson & Nicholls, 2013). This configuration facilitates direct hydride transfer from C4 of the nicotinamide ring to the N5 of FMN, avoiding the high-energy nicotinamide free radical (Berrisford & Sazanov, 2009). The redox potential of NAD^+ is slightly higher than that of the flavin (-0.34 V and -0.38 V at pH 7.5 respectively) and so NADH oxidation is reversible. The FMN bound to bovine complex I has a lowered redox potential, as the oxidised flavin is more strongly bound than its reduced partner (Hirst, 2013). k_{cat} for NADH oxidation by FMN is >15000 s^{-1} and k_{cat}/K_M is ~10^8 $M^{-1}s^{-1}$, which approaches the diffusion limit. The FMN site can also react with synthetic reagents, including $Fe(CN)_6^{3-}$ at very high rates, and transhydrogenation between NADH and an acceptor such as 3-acetylpyridine-adenine dinucleotide ($APAD^+$) can also be catalysed by FMN, although on a slower timescale (Birrell *et al*., 2009).

1.5.2.2. Iron sulphur clusters and intramolecular electron transfer

Seven FeS clusters form a set of stepping-stones to pass electrons from reduced FMN to bound ubiquinone. The isolated cluster 2Fe[24] does not participate and its role is unclear – it has only been observed to accept electrons in the *E. coli* enzyme (Uhlmann & Friedrich, 2005). Electron transfer is considered non-rate limiting, with the rate of transfer between the clusters determined by distance and redox potential. The properties of the FeS chain have predominantly been studied by electron paramagnetic resonance (EPR) and spectra accurately assigned to particular clusters in 2010. Hirst and co-workers showed that the N4 signal is derived from 4Fe[TY]1, N1b from 2Fe[75], N2 from 4Fe[PS] and N3 from 4Fe[51], after previous ambiguities. Consequently it was found that the iron sulphur clusters have alternating redox potentials – a characteristic of many enzymes with sequential cofactors (Roessler *et al.*, 2010). Figure 1.15 shows this arrangement, which is expected to be more kinetically favourable than two adjacent clusters of similar potential (Hirst, 2013). This pattern may also reflect electrostatic interactions between the clusters, which are not accounted for in the experimental profiles (Bridges *et al*., 2012).

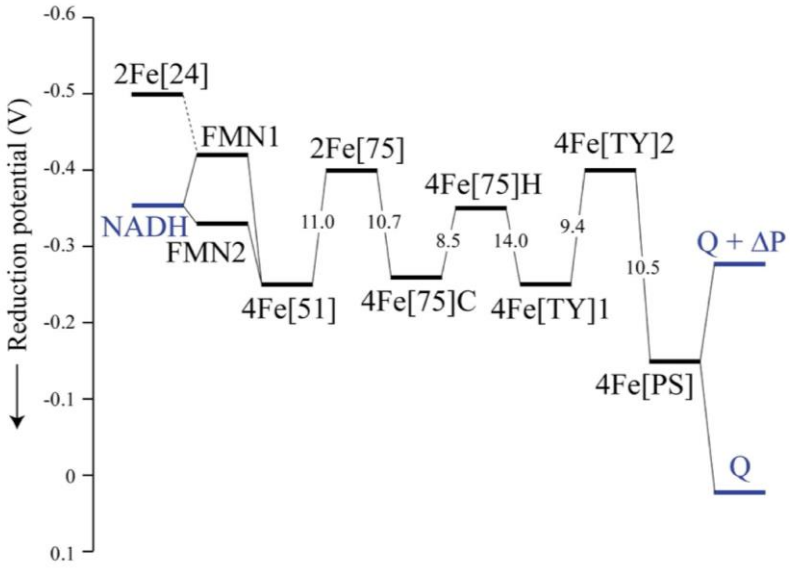


Figure 1.15 Profile of redox potentials of the cofactors of complex I. Based on data from *B. taurus* with substrate potentials shown as two electron potentials. FMN1 denotes the oxidised/semi-reduced couple and FMN2 the semi-reduced/reduced pair. Distances between iron-sulphur clusters are shown in Å as edge-to-edge distances. Two values for Q are ± Δp of 0.15 V. Figure adapted from Hirst, 2009.

1.5.2.3. The ubiquinone binding site

The ubiquinone binding site (Q-site) has been identified in all structures to date; forming a long narrow tunnel that extents ~ 20 Å up from the membrane interface, where the quinone head-group is proposed to bind, although this is not proven (Fedor *et al*., 2017). Both mutagenesis in *E. coli* and *Y. lipolytica* and densities observed in the *T. thermophilus* complex I structure indicate that the head-group binds

through hydrogen bonds to H59 and Y108 of NDUFS2 (Sinha *et al*., 2015; Tocilescu *et al*., 2010; Baradaran *et al*., 2013), placing ubiquinone within 12 Å of the final FeS cluster, N2, to facilitate electron transfer. Overall the tunnel is long enough to enclose most of the 50 Å isoprenoid tail of Q_{10}, although many diverse inhibitors are also suggested to bind in the channel (Murai & Miyoshi, 2016).

Quinone reduction proceeds in two steps, due to single electron transfer from N2. Therefore, a paramagnetic semiquinone intermediate is formed, which is detectable by EPR (Hirst, 2013). Two such intermediates have been described – a fast relaxing and slow relaxing species – that may indicate two quinone binding sites, although this has been excluded experimentally and is not supported by structural data (Ohnishi *et al*., 2010; Hirst & Roessler, 2016). Therefore, the exact nature of these semiquinone intermediates remains unclear. Interestingly the chain length of the quinone substrate has been shown to influence turnover of complex I, with shorter isoprenoid chains correlating with decreased k_{cat} (Fedor *et al*., 2017). This would suggest that the rate limiting step in quinone reduction has slowed or changed and indicates that the rates of binding and dissociation are dictated by tail length. Higher rates with longer isoprenoid chains imply that the tail acts as an anchor and can be guided up the tunnel to correctly position the head group in the binding pocket. Q_1 and Q_2 with short tails may therefore lack this directionality due to increased conformational mobility. Additionally, the end of Q_{10} extends back into the membrane, which may encourage dissociation, whereas with an incompletely filled binding channel other molecules may enter and sit behind smaller quinones, impeding their exit (Fedor *et al*., 2017). There are still a number of outstanding questions regarding the quinone-binding site, particularly as isoprenoids 4-7 lie within a highly charged region of the channel; expected to fill with water molecules. This observation has been supported by molecular dynamic simulations of the *T. thermophilus* structure; although higher resolution structures are needed to further probe this unusual feature (Di Luca *et al*., 2017).

1.5.2.4. Proton pumping

Complex I couples the reduction of ubiquinone to the pumping of protons into the IMS and so contributes to maintaining Δp. An important aspect of complex I function is the stoichiometry of proton translocation, which has been historically considered to be four, with four proton channels proposed (Wikström, 1984; Galkin, 2006). Three channels are found in the antiporter-like subunits ND2, ND4 and ND5, which share homology with Na^+/H^+ Mrp antiporters (Efremov & Sazanov, 2011). The fourth channel was initially identified in the *T. thermophilus* structure at the interface of ND1, ND4L and ND6, and named the E-channel (Baradaran *et al*., 2013). These four channels contain broken transmembrane helices and conformational changes instigated by reduction of quinone at the Q-site may allow complex I to pump protons into the IMS. As described previously, a network of conserved charged residues running along the midline of the membrane domain has been proposed to connect the spatially separated Q-site to these proton channels, thereby transducing energy to couple proton pumping to electron

transfer. The stoichiometry of proton pumping was revisited by Jones *et al.* and used inverted membrane vesicles from *P. denitrificans* and *B. taurus* to confirm earlier measurements of a 4 H^+/2 e^- ratio rather than the more recently proposed 3 H^+/2 e^- ratio by Wikström and Hummer (Jones *et al.*, 2017; Wikström & Hummer, 2012).

The exact coupling mechanism is still uncertain, but two mechanisms have been proposed: one or two-stroke. One-stroke dictates that all energy released from electron transfer from NADH is transduced in a single step with concomitant pumping of protons, which could be feasible due to the four proposed proton channels (Efremov & Sazanov 2012). The Q-site is suggested to be linked to a glutamate-aspartate quartet in the centre of the E-channel by a hydrophilic funnel. The anionic quinol interacts with these negatively charged residues and drives conformational changes in the E-channel that are propagated to the antiporter-like subunits further along the membrane arm. This would drive changes in the pK_a and solvent exposure of key residues, resulting in the translocation of four protons into the IMS or periplasm, Figure 1.16.

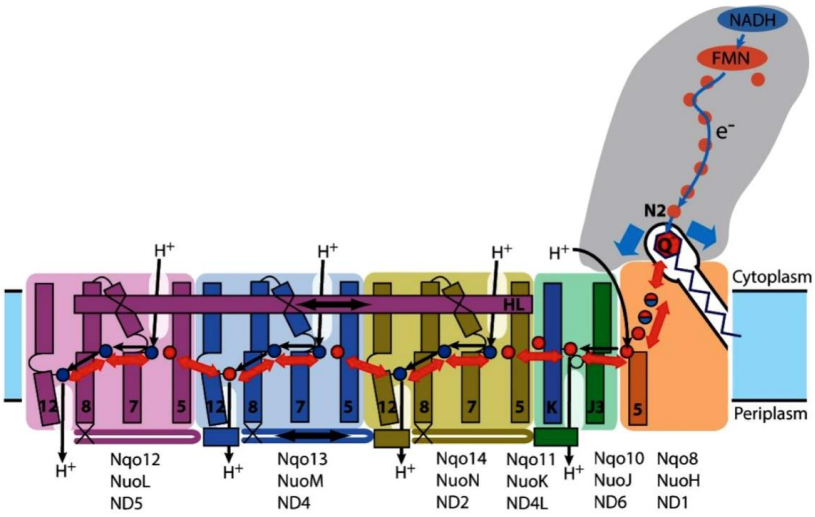


Figure 1.16 Proposed one-stroke mechanism for redox coupled proton transfer. Overview showing that upon electron transfer from N2 the negative quinone initiates a series of conformational changes along the central axis (red arrows) comprising charged or polar residues - shown as circles in blue (lysine or histidine) and red (glutamate). Figure adapted from Sazanov, 2014.

In the two-stroke mechanism energy transduction occurs over two steps, with the formation of a high-energy anionic semiquinone intermediate ($Q^{\bullet-}$) and then an anionic quinol (QH^-) (Brandt, 2011). These intermediates switch between two states – an electron transfer state (E) and a substrate protonation state (P) linked to conformational changes that initiate proton translocation, as shown in Figure 1.17. During the E state the first electron from N2 reduces ubiquinone to $Q^{\bullet-}$, its stabilisation provides the energy required to pump two protons. Then in the P state $Q^{\bullet-}$ is protonated to form $QH^{\bullet}$ and the conformation returns to the E state. A second electron transfer from N2 reduces the protonated semiquinone to QH^- and its stabilisation induces the second stroke for proton pumping. Brandt hypothesises that the same two proton-pumping sites would be used twice, although this is not strictly necessary and the presence of four channels indicates different combinations could be used.

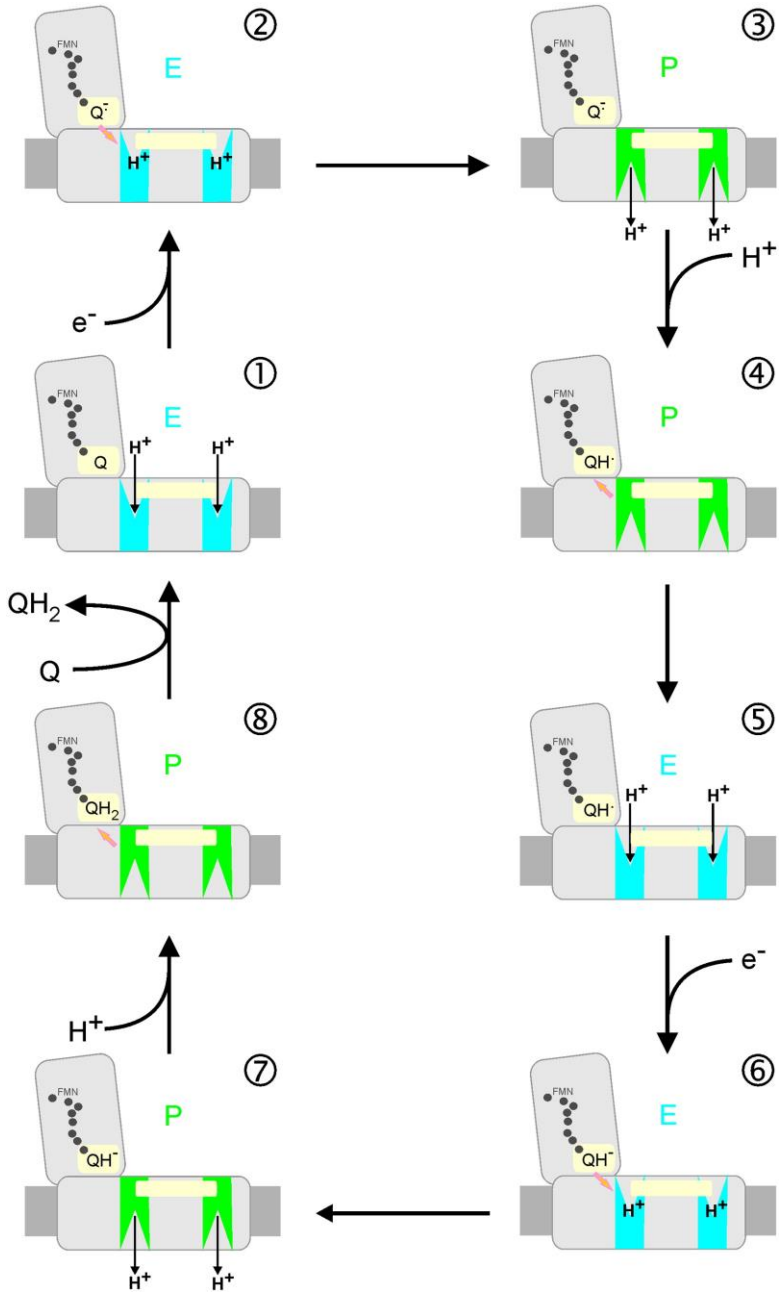


Figure 1.17 Proposed two-stroke mechanism for redox coupled proton transfer. Redox intermediates of ubiquinone drive conformational changes in the membrane arm to pump two protons. Stabilisation of the anionic intermediates $Q^{\bullet -}$ and QH^- cause E (blue) to P (green) state changes. The yellow bar denotes a transmission element between the proton pump sites. Figure adapted from Brandt, 2011.

There has been difficulty ascertaining the route of energy transduction to the proton pumping domains, although increasingly detailed structural data is helping to clarify this. It is thought to involve changes in the Q-binding cleft and propagation into the membrane domain by the chain of charged residues running along its midline, although this is still hypothetical (Zhu *et al*., 2016).

1.6. Mitochondrial disorders and complex I deficiency

Mitochondria are central to regulating cellular bioenergetics; therefore, their dysfunction causes a spectrum of disorders, known collectively as mitochondrial diseases. These are a diverse range of disorders with clinical symptoms presenting from early childhood to late adulthood that arise from dysfunction in the ETC (Swalwell *et al*., 2011). Prevalence is around 1 in 5,000 births and diseases may affect a single organ (the eye in Leber's hereditary optic neuropathy, LHON) or multiple organs (Hoefs *et al*., 2012). Mitochondrial disorders are defined as primary – caused by a mutation in a mitochondrial protein – or secondary – resulting from an external influence on the mitochondria (Koopman *et al*., 2016). Mutations may originate from nuclear DNA or mtDNA due to the dual genomic control of the respiratory complexes (except complex II). nDNA mutations may inherited as autosomal dominant or recessive alleles, with a small number of X-linked mutations. Importantly, mitochondrial disease phenotypes are not only caused by mutations in respiratory chain enzymes but may also result from disruption of essential organellar processes, such as mtDNA maintenance, translation and mitochondrial homeostasis. Defective mtDNA replication caused by mutations in catalytic subunit of DNA polymerase γ (POLG) is responsible for a large proportion of mitochondrial disease phenotypes, which span five different syndromes (Stumpf *et al*., 2013). mtDNA mutations are maternally inherited and heteroplasmy occurs in mtDNA due to copy number, therefore mutations must accumulate and exceed a threshold level for pathogenicity (Schon *et al*., 1997). Therefore, the same mutation can cause different clinical outcomes due to load. For example, the m.8993T>G mutation in MT-ATP6 causes Leigh syndrome with high mutational load and neurogenic weakness with ataxia and retinitis (NARP) with lower heteroplasmy (Uziel *et al*., 1997). Interestingly the most prevalent mtDNA mutations, responsible for ~50% of mtDNA-associated mitochondrial disorders, are found in the 22 tRNA encoding genes, despite these comprising only 5% of the coding sequence in mtDNA (Elson *et al*., 2009). Mutations may affect the structure of tRNAs or prevent post-transcriptional modifications that are necessary for their correct aminoacylation, thus tRNA mutations may severely hamper the translation of mitochondrial-encoded proteins (Yarham *et al*., 2010).

Complex I deficiency is the most common respiratory chain disorder, with an estimated 20-25% caused by nDNA mutations and 25-30% in mtDNA, including both mutations in its subunits, assembly factors and those affecting POLG and tRNAs, particularly $tRNA^{Leu(UUR)}$ (Swalwell *et al*., 2011). Yet, strikingly, nearly half of cases have an unknown genetic basis. Complex I deficiency is one of the most prevalent types in childhood, making up a third of cases due to its early onset, often with severe lactic acidosis and fatality within one year (Fassone and Rahman, 2012). Since the ground-breaking publication by Wallace and co-workers in 1988, which identified a point mutation in ND4 (Wallace *et al*., 1988), complex I mutations have been characterised in different subunits and more recently in assembly factors, although there are still many outstanding questions. This is exemplified by LHON, which is the most common disease caused by mtDNA mutations in complex I subunits but it is late onset, suggesting that

age dependent mtDNA mutation accumulation may dictate disease manifestation (Scheffler, 2015). Figure 1.18A illustrates data from a review of 172 patients with nDNA-encoded mutations (117 in nuclear-encoded subunits and 55 in assembly factors), showing the low survival rates for subunit (core and supernumerary) mutations compared to assembly factor mutations, which display decreased severity. Figure 1.18B highlights the wide range of clinical phenotypes, with more than a third of patients showing Leigh syndrome symptoms and 11% suffering from fatal lactic acidosis as a result of complex I dysfunction (Fassone & Rahman, 2012).

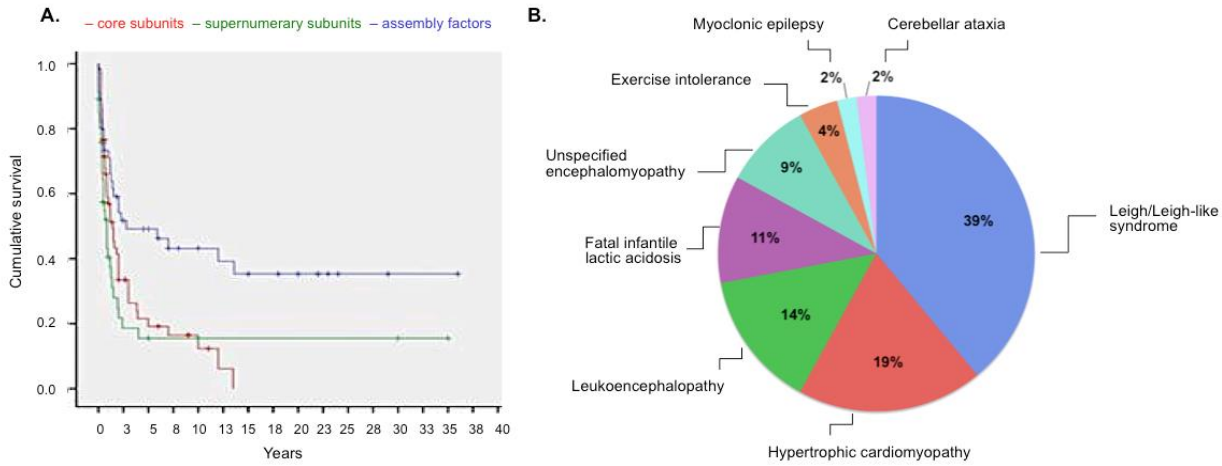


Figure 1.18 Complex I deficiency caused by nuclear-encoded DNA mutations. A. Kaplan-Meier survival curves for 172 patients with mutations in different proteins related to complex I. Red shows nuclear encoded core subunits, green shows supernumerary nDNA encoded subunits and blue assembly factors. **B. Clinical phenotypes of patients suffering from complex I deficiency.** The pie chart shows the diverse range of symptoms caused by complex I subunits and assembly factor mutations in the 172 patients. Figure adapted from Fassone & Rahman, 2012.

Complex I disorders are usually diagnosed by rotenone-sensitive NADH:ubiquinone oxidoreductase activity in muscle biopsies, however caveats remain on the reference activity used (Liolitsa *et al*., 2000). However, other methods are emerging: using oxygen consumption to assess the respiratory capacity of mitochondria, immunocapture and Blue Native polyacrylamide gel electrophoresis (BN-PAGE) to address subunit composition and structural integrity, along with in-gel activity assays (Fassone & Rahman, 2012). Additionally, there are some secondary complex I deficiencies, including those related to Parkinson's and thyroid cancers (Schapira *et al*., 1990; Gasparre *et al*., 1997). There is no cure for complex I deficiency, and most treatments address the symptoms only, although some patients exhibit riboflavin responsiveness due to the importance of this vitamin in flavin cofactor biosynthesis - the role of riboflavin will be further discussed in Chapter 3 (Bersen, 1993). Targeting reactive oxygen species (ROS) production and mitobiogenesis are also approaches that may be viable treatment options for complex I disorders, but these are all in the early stages of development, with no drugs in clinical trial (Rahman, 2015). The genetic cause of complex I deficiencies is becoming increasingly well understood and documented but defining the mechanisms underpinning these defects is trailing behind. This is in part due to the limitations in sample amounts obtained from biopsies and that cultured patient fibroblasts

do not always mirror the defects observed (Ruitenbeek *et al*., 1996). Therefore, a reliable model system is crucial for investigating complex I defects. There has been development of mouse models for specific nDNA mutations, but these do not perfectly replicate the human disease. For example, although the NDUFS6 knockdown mouse, generated by gene trap (Ke *et al*., 2012), displays a complex I deficiency with isolated hypertrophic cardiomyopathy (HCM), there have been no human patients with NDUFS6 mutations causing HCM reported (Fassone & Rahman, 2012). Human cells in culture also allow for genetic manipulation, however targeted mutagenesis of mtDNA remains a technical challenge and precludes the study of mutations in the core mtDNA-encoded subunits of complex I (Irwin *et al*., 2013). The prognosis for complex I deficiency is very poor – as shown by survival rates in Figure 1.18A - and due to the complexity and size of this enzyme the importance of understanding its mechanism and structure are fundamental to addressing the effects of mutations; with the goal to find a therapeutic intervention that may be used to restore mitochondrial respiratory activity to patients suffering from these severe disorders.

1.7. Aims of the thesis

The work in this thesis focuses on a number of aspects of complex I structure and function. Flavin handling by the cell was explored to investigate when the FMN cofactor is incorporated into complex I. Additionally, the quinone-binding site of complex I was studied using mutagenesis in the *Y. lipolytica* model system to affect quinone binding and catalysis. Variants in human complex I genes were also analysed by their location in a human model structure of the enzyme.

- First, Chapter 3 aims to explore when FMN is inserted into the flavin binding subunit (NDUFV1) of complex I during assembly, by manipulating flavin handling in cultured human cells.
- Chapter 4 aims to determine the physiological role of the mitochondrial carrier protein SLC25A32, due to its putative role as a flavin transporter, and by knocking out the protein probe the affect of disrupted flavin transport on complex I activity and structure.
- Chapter 5 aims to investigate the role of specific residues in the quinone channel, by mutagenesis of the NDUFS2 subunit in *Y. lipolytica* and incorporation of variant complex I into proteoliposomes for kinetic study.
- Finally, Chapter 6 aims to combine clinically-relevant variants with the most detailed information on complex I structure – using a human complex I model to identify which characteristics can be used to predict the likelihood that a variant will be pathogenic.

Chapter 2

Materials and Methods

2.1. Materials

All materials were purchased from Sigma-Aldrich or Thermo Fisher Scientific unless otherwise stated.

2.2. Cell Culture

143B cells (143B CyWT) are immortalised human osterocarcoma patient cybrids – haplotype H13b1b – prepared by Zeviani and co-workers as previously described (Corona *et al*., 2001; King and Attardi, 1996). 143B CyWT cells were cultured at 37 ºC in high glucose Dulbecco's modified eagle medium (DMEM, Life Technologies) + 10% foetal bovine serum (FBS, labtech.com) under 5% (v/v) CO_2. DMEM contained 1 μM riboflavin (data from supplier) and 10% FBS contributed another 0.77 nM (Said *et al*., 1998).

HAP1 cells**,** derived from male chronic myelogenous leukemia cell line KBM-7 (Horizon Discovery), were cultured at 37 ºC in Iscove's Modified Dulbecco's Medium (IMDM, Life Technologies) + 10% FBS under 5% v/v CO_2. IMDM contained 1 μM riboflavin (data from supplier) and 10% FBS contributed another 0.77 nM (Said *et al*., 1998).

Riboflavin depletion of mammalian cells in culture

Riboflavin-free DMEM (Life Technologies) was used to condition cells: the riboflavin content was adjusted with different proportions of riboflavin-free DMEM and standard DMEM or IMDM (for 143B and HAP1 respectively) in 25% steps with two passages at each stage. All media contained 10% FBS.

The inhibitor lumiflavine (Santa Cruz) has been shown to inhibit riboflavin uptake by cultured cells (Said *et al.*, 1998). 1 mM lumiflavine in Dulbecco's phosphate buffered saline (DPBS, Life Technologies) was filter sterilised and stored at -20 ºC. 1 μM, 5 μM and 10 μM concentrations were used by addition to DMEM + 10% FBS and incubated with 143B cells for 72 hours.

To observe changes in growth rate and cellular morphology of treated 143Bs, an Incucyte HD was used to periodically record their physical characteristics over 7 days. Cells were seeded into 6 well plates at 25,000 cells/well (which allowed 143B CyWT to reach confluence after one week, HAP1 cells were plated at 14,000 cells/well). Each well was scanned at 16 positions with images taken every 2 hours, resulting in an image bank for each growth condition.

Following conditioning, cells were maintained in low riboflavin conditions of riboflavin-free DMEM + 10% FBS (total 0.77 nM riboflavin) for one month before further investigations were undertaken - these cells are hereafter referred to as 'conditioned'.

2.3. Genetic manipulation of protein expression in cultured human cells

2.3.1. siRNA knockdown in 143B CyWT cells

Transcripts of target proteins were transiently suppressed by 50 nM or 100 nM siRNA. 143B CyWT cells were plated at a density of 75,000 cells/well in 6 well plates and after 24 hours siRNA was added using Lipofectamine™ 2000 and OptiMEM (both Life Technologies). All siRNA were purchased from Eurofins Genomics and three different oligonucleotides were designed for each protein (Table 2.1), using the Eurofins online tool. Knockdown cells were incubated for 72 hours to maximise changes in protein expression.

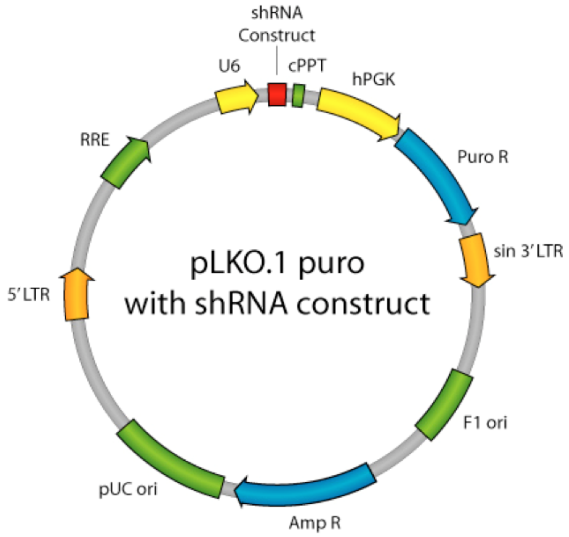
Protein	siRNA	Target sequence
SLC25A32	1	CAACCAGCATATCAATAGA
	2	GGATGCCAGAATACTAGTA
	3	GTATGATGCTGTTGTTAAC
RFK	1	CGACTAGAGTTACCAGAAC
	2	GCTGATATATCCACTGGTA
	3	TACGAAGAAGTCTATGGAA
FADS1	1	AGGAGTCGCTTGTCAAGGA
	2	GGATCTGGTTAGAGAAGAC
	3	GTGGTAGGTTCTCAAGAGA

Table 2.1 siRNA targets for SLC25A32, RFK and FADS1. siRNA chosen for the mitochondrial isoform of FLAD (FADS1) only targeted sequence unique to this isoform, to leave the cytoplasmic FADS2 unaffected. siRNA were synthesised by Eurofins Genomics.

2.3.2. Stable knockdown in 143B CyWT cells using shRNA

For stable expression of shRNA in human cell lines the lentiviral vector pLKO.1 was used (Figure 2.1), in conjunction with a second-generation lentiviral system comprising packaging and envelope plasmids (discussed below).

A.



B.

Element	Description
U6	Human U6 promoter drives RNA Polymerase III transcription for generation of shRNA transcript
cPPT	Central polypurine tract improves transduction efficiency of lentivirus by facilitating import of the vector pre-integration complex in the transduced cell nucleus.
hPGK	Human phosphoglycerate kinase promoter drives expression of puromycin resistance gene
Puro R	Puromycin resistance gene for selection of pLKO.1 plasmid in mammalian cells
sin 3'LTR	3' Self-inactivating long terminal repeat
f1 ori	f1 bacterial origin of replication
Amp R	Ampicillin resistance gene for selection of pLKO.1 plasmid in bacterial cells
pUC ori	pUC bacterial origin of replication
5'LTR	5' long terminal repeat
RRE	Rev response element

Figure 2.1A. pLKO.1 vector map. pLKO.1 vector containing an shRNA insert (red), showing key features of the vector for use in shRNA expression. **B. Table of key features.** The pLKO.1 vector contains a number of engineered features chosen to improve shRNA expression within the target cell, resistance genes for bacteria and mammalian selection and a self-inactivating (SIN) device for safety. A deletion in the 3'LTR of the viral genome that is transferred into the 5'LTR after one round of reverse transcription abolishes transcription of the full-length virus after it has incorporated into a host cell. Figure A. from Moffat *et al.*, 2006.

The pLKO.1 vector was the backbone used in Sigma MISSION® shRNA and also used later for shRNA cloning. It contains an ampicillin resistance gene for bacterial selection and a puromycin resistance gene for selection within mammalian cell lines (Figure 2.1B).

A. Forward shRNA oligonucleotide

5' CCGG–21bp sense–CTCGAG–21bp antisense–TTTTTG 3'

Reverse shRNA oligonucleotide

5' AATTCAAAAA–21bp sense–CTCGAG–21bp antisense 3'

B.

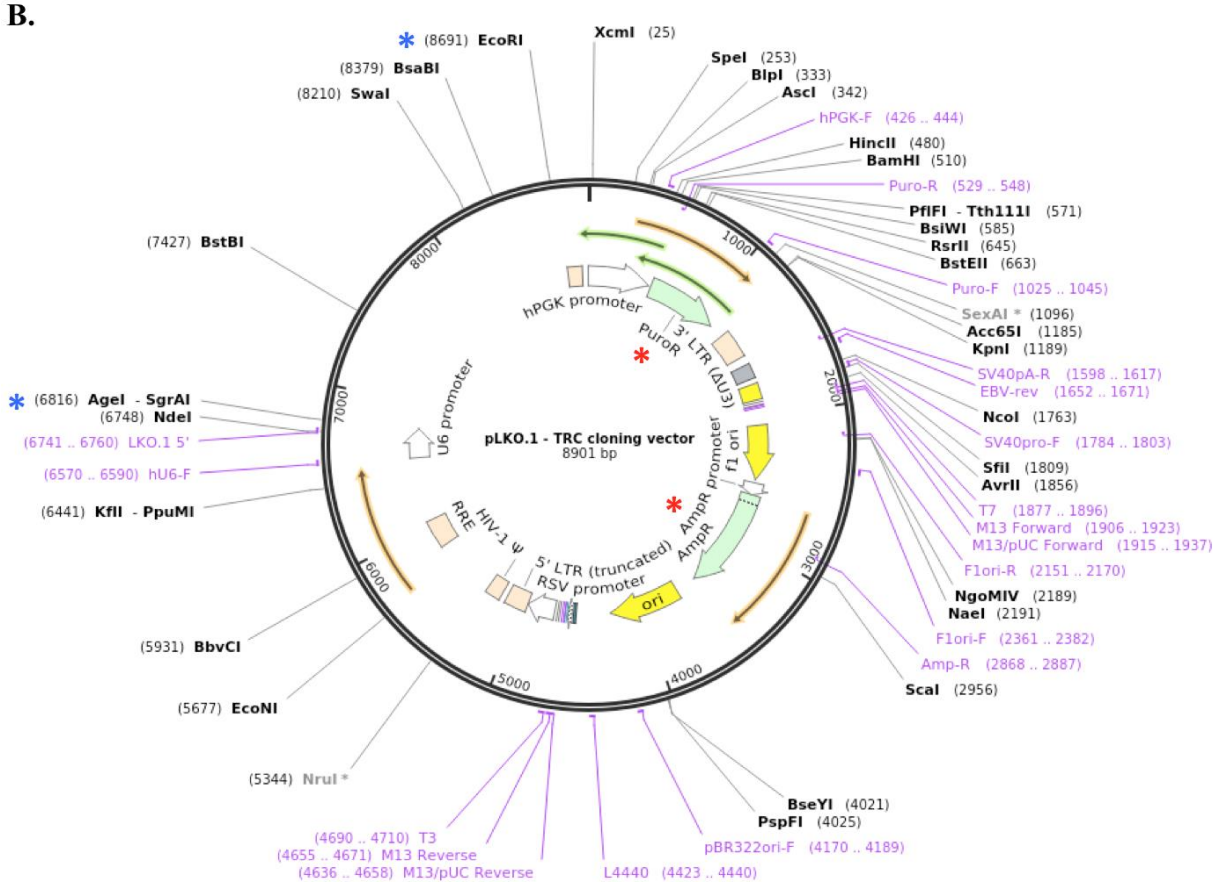


Figure 2.2A. Template for forward and reverse oligonucleotides containing the shRNA target sequence chosen for the specific gene of interest. This template was used by Sigma MISSION® shRNA vectors and for construction of pLKO.1 shRNA vectors. For shRNA cloning the Addgene protocol was used - two oligonucleotides were designed as shown and annealed to produce dsDNA with overhangs for ligation into the restricted pLKO.1 vector (red for AgeI and dark blue for EcoRI). The template would generate a hairpin structure when transcribed that could be processed into siRNA within the cell to promote gene silencing. **B. Schematic of pLKO.1 vector.** The main features of the vector are highlighted. Restriction at EcoRI and AgeI sites (blue asterisks) decreased the vector size to 7 kb and the two resistance genes (red asterisks) allowed selection in bacterial and mammalian systems. The plasmid (Addgene #10878) was created by Root and co-workers (Moffat *et al.*, 2006). Plasmid map is available from https://www.addgene.org/10878/.

For each target gene two Sigma MISSION® shRNA pLKO.1 vectors were chosen for lentiviral transfection of 143B CyWT cells. The vector contained an insert made by annealing two shRNA containing oligonucleotides, templates shown in Figure 2.2A - the shRNA sequence was inserted as the sense and antisense strands between conserved short base sequences. This sequence then formed a hairpin on transcription due to annealing of complementary sequences and the shRNA was processed within the cell to form siRNA, to target its complementary mRNA sequence for expression silencing.

Sigma MISSION® shRNA pLKO.1 vectors	
Construct	**Sequence**
SLC25A32	
TRCN000043090	CCGGGCAGCAACATACCCATATCAACTCGAGTTGATATGGGTATGTTGCTGCTTTTTG
TRCN000043088	CCGGGCCCTTCAGTTTATGGCATATCTCGAGATATGCCATAAACTGAAGGGCTTTTTG
FADS1	
TRCN0000300338	CCGGGCACTGGTCAACCTTTCAGATCTCGAGATCTGAAAGGTTGACCAGTGCTTTTTG
TRCN0000294377	CCGGGGCTGGGAGAAGCTATCATTGCTCGAGCAATGATAGCTTCTCCCAGCCTTTTTG
RFK	
TRCN000037599	CCGGCCCATATTACAAGAATACGAACTCGAGTTCGTATTCTTGTAATATGGGTTTTTG
TRCN000037602	CCGGGCATACCTTCAAAGAGGACTTCTCGAGAAGTCCTCTTTGAAGGTATGCTTTTTG

Table 2.2 Sigma MISSION® shRNA pLKO.1 constructs chosen for SLC25A32, FADS1 and RFK stable knockdown. Two constructs for each protein target were chosen, with care taken with FADS1 to target only this isoform of the FLAD1 gene. Forward shRNA oligonucleotide sequences are shown and coloured to match Figure 2.2A.

Design of non-commercial shRNA constructs

Using the online tool from the Whitehead Institute (Yuan *et al*., 2004) the following guidelines aided the design of shRNA for the genes SLC25A32 and FADS1, which could not be knocked down using commercial shRNA vectors (section 3.6.1.1).

- 21 nucleotides with pattern AA(N_{19}) or NAR(N_{17})YNN (N is any nucleotide, R is A or G and Y is C or T) chosen when possible
- G-C content of 36-52%
- No stretches of four or more nucleotide repeats (TTTT is a termination signal)

Multiple target sequences were chosen for each gene and covered different locations (Table 2.3). All sequences were checked for off-target interactions using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et* al., 1990).

FADS1		
shRNA 1	*sense*	5'GCTCAGTAATCTGAAGCTTGG 3'
	antisense	3'CGAGTCATTAGACTTCGAACC 5'
shRNA 2	*sense*	5'GGGAACACGTTTATTCCAGAG 3'
	antisense	3'CCCTTGTGCAAATAAGGTCTC 5'
shRNA 3	*sense*	5'GGAGTCGCTTGTCAAGGATCT 3'
	antisense	3'CCTCAGCGAACAGTTCCTAGA 5'

SLC25A32		
shRNA 1	*sense*	5'AAGTGAACTGCAGTTGATATT 3'
	antisense	3'GGCACTTGACGTCAACTATAA 5'
shRNA 2	*sense*	5'ATCCTAGATCCAGTTCTTGTT 3'
	antisense	3'TAGGATCTAGGTCAAGAACAA 5'
shRNA 3	*sense*	5'ACTGACACGAAGGATAGAGTT 3'
	antisense	3'TGACTGTGCTTCCTATCTCAA 5'

Table 2.3 shRNA sequences chosen for FADS1 and SLC25A32 knockdown. Three shRNA sequences were chosen for each gene, covering different regions. The template in Figure 2.2.A was then used to design the shRNA hairpin oligonucleotides for insertion into the pLKO.1 vector. Primers were synthesised by Sigma.

Initially, to insert the shRNA template into the multiple cloning site (MCS) the restriction digest method outlined by Addgene was employed (http://www.addgene.org/tools/protocols/plko/#B). The pLKO.1 vector was restricted with EcoRI and AgeI (New England Biolabs, NEB), removing a 1.9 kb stuffer fragment from the MCS. Two complementary oligonucleotides were designed for each target sequence (as shown in Figure 2.2A), which would anneal with overhangs allowing ligation into the linearised vector. The oligonucleotides were annealed to form dsDNA and then ligated into pLKO.1 with T4 DNA ligase (NEB). After transformation into D5-α competent *E. coli*, colonies were assessed by mini prep (Qiagen). However no positive colonies could be obtained by this method, despite many repetitions, protocol alterations and sequencing of the pLKO.1 vector. A different approach was therefore tested.

pLKO.1 was amplified with primers that also contained part of the insert and so after a PCR cycle the product was a linearised vector with the insert present, which could then be ligated. To achieve the same result as the restriction digest approach, primers were designed for the desired end construct with the shRNA template present between the AgeI and EcoRI sites, Figure 2.2A. 20 bases of pLKO.1 either side of the insert were used in the primers and, as shown in Figure 2.3A, the 58 base insert was split in half - the forward primer comprised the last 29 bases of the insert at the 5' (part 2.) followed by the 20 nucleotides downstream of the insert in pLKO.1 (part 3.). The reverse primer included the first 29 bases of the insert (part 1.) and the 20 bases that precede it in pLKO.1 (parts 4.). Figure 2.3A shows primer annealing and extension during PCR by the two primers, which were 5' phosphorylated to allow for

effective ligation. Figure 2.3B shows the PCR product generated, which was then ligated to form the vector containing the desired shRNA sequence – example SLC25A32 shRNA 1.

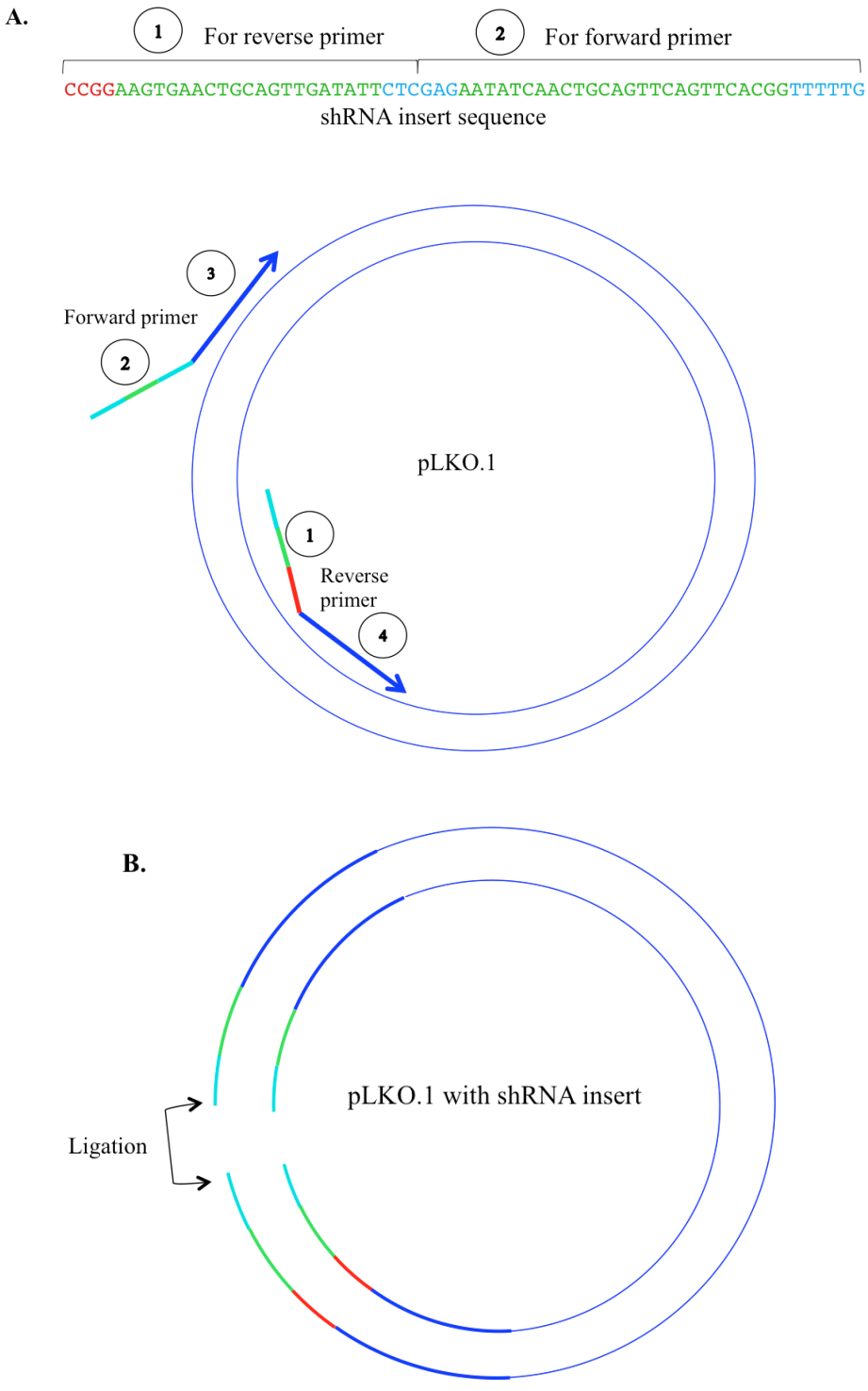


Figure 2.3 Cloning of pLKO.1 to contain shRNA inserts. A. Primer annealing and extension of the vector to insert the shRNA sequence into the MCS and remove the stuffer fragment. The shRNA sequence (example: SLC25A32 shRNA 1) was split in two with half used in the forward and reverse primers (1 and 2). This was attached to 20 bases upstream or downstream of the insert site to make the full primers (3 and 4, dark blue lines). **B. PCR product** of linearised pLKO.1 without the stuffer fragment between the AgeI and EcoRI restriction sites and containing the shRNA sequence, in two halves. Ligation by T4 DNA ligase was used to re-circularise the construct and produce the complete vector.

FADS1		
shRNA 1	*forward primer*	5'[Phos]GAGCCAAGCTTCAGATTACTGAGCTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGCCAAGCTTCAGATTACTGAGCCCGGTGTTTCGTCCTTTCCACAAG 3'
shRNA 2	*forward primer*	5'[Phos]GAGCTCTGGAATAAACGTGTTCCCTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGCTCTGGAATAAACGTGTTCCCCCGGTGTTTCGTCCTTTCCACAAG 3'
shRNA 3	*forward primer*	5'[Phos]GAGAGATCCTTGACAAGCGACTCCTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGAGATCCTTGACAAGCGACTCCCCGGTGTTTCGTCCTTTCCACAAG 3'

SLC25A32		
shRNA 1	*forward primer*	5'[Phos]GAGAATATCAACTGCAGTTCACGGTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGAATATCAACTGCAGTTCACTTCCGGTGTTTCGTCCTTTCCACAAG 3'
shRNA 2	*forward primer*	5'[Phos]GAGAACAAGAACTGGATCTAGGATTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGAACAAGAACTGGATCTAGGATCCGGTGTTTCGTCCTTTCCACAAG 3'
shRNA 3	*forward primer*	5'[Phos]GAGAACTCTATCCTTCGTGTCAGTTTTTTGAATTCTCGACCTCGAGACA 3'
	reverse primer	5'[Phos]GAGAACTCTATCCTTCGTGTCAGTCCGGTGTTTCGTCCTTTCCACAAG 3'

Table 2.4 Primers used for amplification of pLKO.1 to include the shRNA insert. To avoid the annealing step primers were designed to amplify the pLKO.1 vector, linearising it and adding half the shRNA template to each end. These ends were then ligated to form the construct containing the shRNA for transformation in D5-α *E. coli.* 5' phosphate was added to aid ligation efficiency of the blunt ended vector. Sequences are coloured to match previous figures. Primers were synthesised by Sigma.

PCR reaction

A thermocycler programme was designed taking into account the different temperatures of annealing for the individual primer pairs and the two stages of annealing, the initial binding of just the regions complementary to the vector (Table 2.4 in navy) and later the full primer binding (Table 2.4 in navy plus green, red and cyan). These were calculated using the NEB T_m calculator for the primer pairs (http://tmcalculator.neb.com/#!/main). For a total volume of 50 µl the following reaction was prepared: 2.5 µl of 10 µM forward primer, 2.5 µl of 10 µM reverse primer, 10 µl Q5® polymerase buffer, 10 µl Q5® enhancer, 20 ng pLKO.1 template, 1 µl 10mM dNTPs and 1µl Q5® polymerase. The PCR thermocycler programme used is shown in Table 2.5.

Temperature (°C)	Duration (min:sec)	Loops
98	00:30	
98	00:10	5 loops
Initial primer binding temperature	00:20	
72	03:30	
98	00:10	25 loops
Full primer binding temperature	04:00	
72	05:00	
4	Hold	

Table 2.5 Template thermocycler programme for PCR reaction to amplify shRNA insert containing pLKO.1. The programme was adapted for the individual primer pairs (initial primer binding and full primer binding temperatures) and the extension set at 3 min 30 sec for a 7 kb construct, 30 sec/kb.

To visualise PCR products 0.8% agarose gels were prepared by dissolving 0.8% (w/v) agarose (BioGene) in 1X Tris-Borate-EDTA buffer (TBE) by heating. The solution (50 ml), with 2 µl of SYBR™ Safe (Thermo Fisher) added was then set in a mold with a sample-loading comb inserted. 5 µl of the amplified DNA product alongside 5 µl MassRuler Express forward DNA ladder (Thermo Fisher) were resolved at 100 V for 40 minutes. SYBR™ Safe that had intercalated with the DNA was imaged on a ChemiDoc™ imaging system (Bio-Rad) to confirm the product formed was ~ 7 kb.

Ligation

The ligation mixture was made up to 20 µl with 2 µl PCR product, 2 µl T4 DNA ligase buffer and 1 µl T4 DNA ligase and incubated at 37 °C for two hours. After heat inactivation at 65 °C for 20 minutes, 1 µl Dpn1 (NEB) was added to digest the methylated pLKO.1 template, incubated for 15 minutes at 37 °C and heat inactivated at 80 °C for 20 minutes.

Transformation and colony analysis

2 µl of ligation mix were added to 50 µl competent D5-α cells and incubated on ice. A 30 second heat shock at 42 °C was followed by 5-minute incubation on ice and addition of 950 µl SOC medium. After 1 hour of incubation at 220 rpm and 37 °C, *E. coli* were plated and incubated overnight at 37 °C on 100 µg/ml ampicillin lysogeny broth (LB) containing agar. Resulting colonies were analysed by mini preparation (QIAprep spin kit) of plasmids, which were analysed by 0.8% agarose gel electrophoresis positive colonies sequenced to confirm successful construct formation (Genwiz).

Lentiviral particle production

To generate stable knockdowns the shRNA construct (commercial or synthesised as above) was introduced into the host cells by lentivirus using a multistep protocol from Addgene undertaken over several days (http://www.addgene.org/tools/protocols/plko/#E). HEK293T cells are highly transfectable human embryonic kidney cells that replicate vectors containing the SV40 region of replication, including pLKO.1 and were used to produce lentiviral particles.

Firstly HEK293T cells were plated at 2.5 $x10^6$ cells per 10cm dish, in DMEM + 10% FBS. After 24 hours the media was substituted for DMEM + 10% FBS + 25 µg/ml chloroquine – to improve transfection efficiency. This was done one hour before transfection to minimise exposure to chloroquine, due to its toxic effects on cells. Lentiviral particles were produced by transfection of HEK293T with the shRNA construct and two additional vectors - psPAX2, packaging vector and pMD2.G, envelope vector – shown in Figure 2.4.

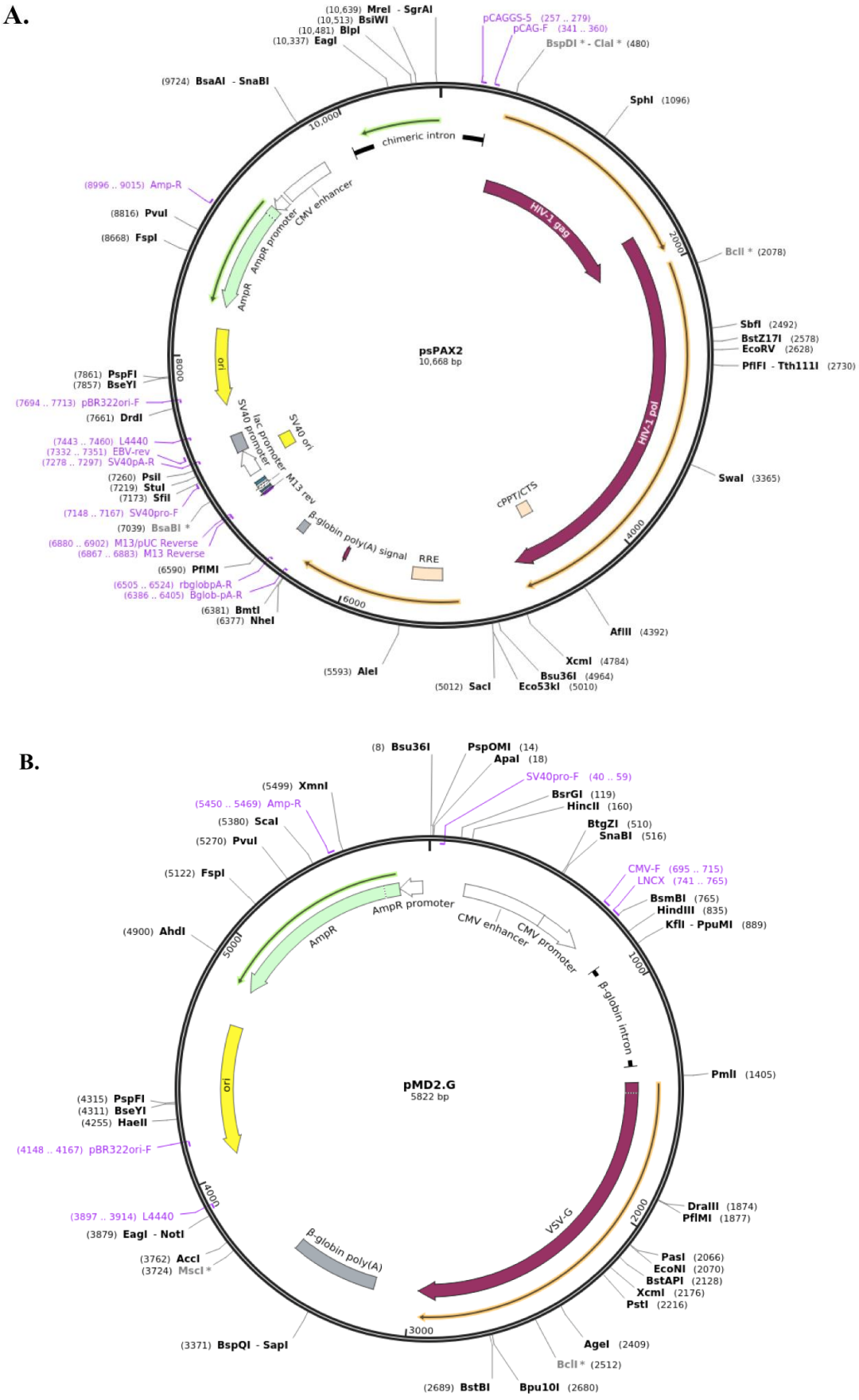


Figure 2.4 Vector maps for envelope and packaging plasmids used for lentiviral production in HEK293T. A. psPAX2 packaging vector (Addgene # 12260). **B. pMD2.G** envelope vector (Addgene #12259). Plasmids were created by Truno and co-workers (unpublished) and maps are available from https://www.addgene.org/12260/ and https://www.addgene.org/12259/, respectively.

psPAX2 and pMD2.G were required for the expression of lentivirus structural proteins in HEK293T cells. These two plasmids were amplified by transformation into D5-α cells and plated onto LB agar containing 100 μg/ml ampicillin. Then small volume liquid cultures were made and plasmid DNA purified by mini prep purification (Qiagen).

The transfection mix comprised the three vectors, serum free DMEM and Fugene HD (Promega). 10.00 μg pLKO.1 shRNA construct, 6.55 μg psPAX2 and 3.5 μg pDM2.G were used for each shRNA construct type and mixed with serum free DMEM to a final volume of 940 μl, then 60 μl of Fugene® HD transfection agent was added and the mixture incubated for 30 minutes, all under sterile conditions. After incubation the transfection mix was added drop-wise to the HEK293T plates with chloroquine containing DMEM and incubated for 8-12 hours at 37 °C, 5% CO_2. The media was then changed, to remove chloroquine and the transfection agent, and replaced with DMEM + 10% FBS, 8 ml/plate, for a 48 hour incubation.

The following day the host cells, 143B CyWT, were plated at 2 x10^6 on 10 cm plates in DMEM + 10% FBS and incubated for 24 hours at 37 °C, 5% CO_2. One dish was plated for each construct and an additional control plate for puromycin selection.

48 hours after transfection, media from the HEK293T cells was collected and contaminating cells removed by centrifuged at 4,100 xg for 5 minutes. The media contained lentiviral particles produced by the HEK293T cells using the transfected plasmids, which was removed and filtered through 0.45 μm filters to obtain ~7.5 ml of particle containing media/plate. 143B cells were then prepared for infection, media was substituted with 8 ml DMEM + 10% FBS + 8 μg/ml polybrene, to aid lentiviral infection. The harvested and filtered lentiviral particle-containing DMEM was added and 143B cells incubated overnight at 37 °C, 5% CO_2.

Due to the presence of the puromycin resistance gene on pLKO.1, 143B cells containing and expressing the plasmid construct should be resistant to 1 μg/ml puromycin, which was added in fresh media after overnight incubation. The cells were then checked by Western Blot for decreased target protein expression after a further 4 days of incubation. Non-infected 143B cells were used as a positive puromycin control; after a few days of puromycin incubation these cells died. Stably knocked down 143Bs were maintained in puromycin containing DMEM to prevent reversion.

2.3.3. Stable knockout using CRISPR

The clustered regularly interspaced short palindromic repeats (CRISPR) technique was employed to produce a more severe phenotype than siRNA or stable knockdowns by shRNA.

The single guide RNA (sgRNA) required for Cas9 recognition of the target gene sequence was designed to include a short sequence homologous to the target and requires a 2-6 nucleotide protospacer motif (PAM) directly downstream of the target sequence. The protocol for designing and testing the Cas/CRISPR system for SLC25A32 sgRNA followed the pipeline described in Figure 2.5 (guidelines from Clontech Guide-it™ sgRNA *in vitro* transcription user manual).

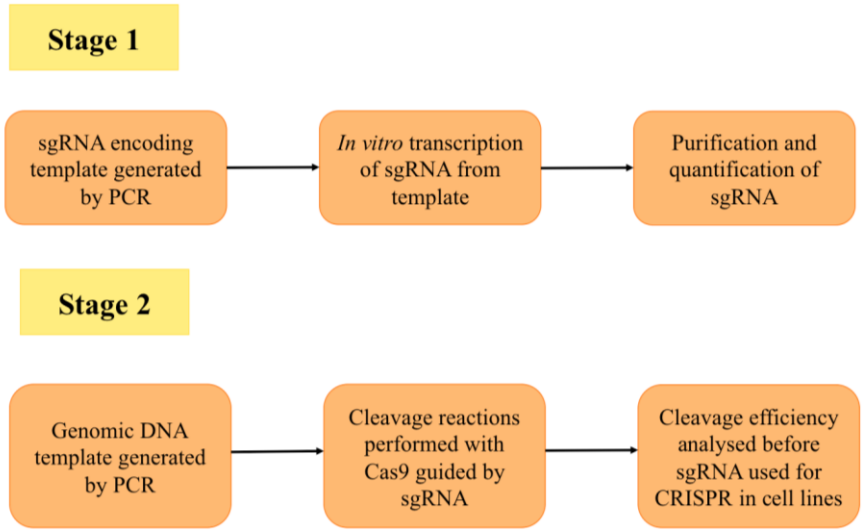


Figure 2.5 Schematic of CRISPR sgRNA design for gene knockout. sgRNA was designed against the target gene using the online tool CHOPCHOP. A DNA template was purified from HAP1 cells and used to test the efficiency of the Cas9 mediated cleavage before the sgRNA was cloned into the pSpCas9(BB)-2A-GFP vector for transformation into HAP1s. Protocol from Clontech Guide-it™ sgRNA *in vitro* transcription user manual.

The sgRNA-encoding DNA template was designed using the online tool CHOPCHOP (Montague *et al*., 2014). The human gene sequence ENSG00000164933 was used in SLC25A32 analysis by CHOPCHOP, using the transcript identification code NM_030780. The most recent version of the annotated human genome (hg38/GRCh38) was used as the *"in"* component and Cas/CRISPR in the *"using"* input, which resulted in an output of different sgRNA sequence targets across the gene, ranked by predicted efficiency. A number of other parameters were provided – including GC content, self-complementarity and predicted off-targets. A traffic light coding system aided selection and only green suggestions were considered. The best three sequences available in exon 1 were for sgRNA-encoding template design. Exon 1 was the best target as a deleterious frame-shift early on in the gene would be more likely to result in complete loss of protein function. Target sequences chosen were all located in exon 1 and received the green coding, with high-predicted efficiency and no predicted off-target effects.

SLC25A32	Sequence	Strand
sgRNA 1	GGTTCTCATACCGGACGTGGCGG	+
sgRNA 2	CGGCGAAGCGGATCTTCACGAGG	+
sgRNA 3	CGGTATGAGAACCTGATAGCGGG	-

Table 2.6 Three sgRNA chosen from CHOPCHOP results for SLC25A32. The top three sgRNA suggested for exon 1 targeting. Highlighted in blue is the PAM sequence, which was omitted from subsequent steps for making the sgRNA template.

Stage 1 – producing the sgRNA template

Figure 2.6 Example of forward primer design. The 20bp sgRNA (without the downsteam PAM) was inserted between a T7 promoter sequence 5' and a 3' Scaffold template specific sequence, to produce a specific forward primer for each sgRNA.

The forward primer (designed as shown in Figure 2.6) was produced by Sigma, diluted to 10 µM and used along with the Guide-it™ sgRNA *in vitro* transcription kit (Clontech) to generate the sgRNA-encoding DNA template. 5 µl Guide-it™ Scaffold template was mixed with 1 µl forward primer and 19 µl RNAse free dH_2O; with the entire 25 µl used to resuspend an aliquot of High Yield PCR EcoDry™ Premix (Clontech). The following thermocycler programme was then used:

95 °C – 1 minute
95 °C – 30 sec
68 °C - 1 minute] x33 cycles
68 °C – 1 minute

5 µl of the product was loaded onto 1.8% agarose gel, along with 5 µl Mass Rule Express Forward DNA Ladder; a single 140 bp band was expected for successful amplification.

In vitro transcription was then carried out to generate RNA from 100 ng of the DNA template generated above. 7 µl transcription buffer and 3 µl T7 polymerase from the Guide-it™ *in vitro* transcription kit (Clontech) were mixed with the DNA template and made up to 20 µl before incubation at 42 °C for an hour.

RNA was purified using DNAse I digestion, phenol:chloroform extraction and ethanol precipitation. 2 µl of RNAse free DNAse I (Clontech) was added to 20 µl of the transcription mix and incubated at 37

°C for an hour, then made up to 100 µl with RNAse free dH_2O. The following steps were carried out in a fume hood – 100 µl phenol:chloroform:isoamylalcohol (25:24:1) was added and the sample vortexed. Centrifugation at 13,700 xg for 2 minutes separated the solvent layers. The top aqueous layer, containing RNA, was retained and the organic lower layer, containing contaminating protein, was discarded. 200 µl of chloroform was added to the upper phase, centrifuged and the aqueous phase retained. 40 µl of 3M sodium acetate was then added with 400 µl isopropanol, vortexed and incubated for 5 minutes at room temperature before centrifugation for 5 minutes at 19,500 xg. The pellet formed was carefully rinsed with 80% ethanol and centrifuged again. The RNA containing pellet was air-dried for 15 minutes before being resuspended in 20 µl RNAse free dH_2O. Aliquots were then snap frozen in liquid N_2 and stored at -80 °C until further use.

Stage 2: Testing Cas90/sgRNA cleavage efficiency of target gene sequences

A genomic DNA template (gDNA) was required to test sgRNA cleavage specificity, before being used in the chosen cell line – HAP1. gDNA PCR primers were designed around the target sequence to generate a ~500 bp template, which when cut would produced two fragments of different sizes, to allow visualisation of cleavage.

gDNA was first prepared from HAP1 cells (1-2 $x10^6$ cells), lysed with 300 µl Cell Lysis Solution (Qiagen) and protein precipitated with 100 µl Protein Precipitation Buffer (Qiagen). After incubation on ice for 5 minutes the contaminating protein was removed by centrifugation (16,000 xg) and 300 µl isopropanol added to the supernatant. After thorough mixing by inversion the sample was centrifuged and the white DNA-containing pellet washed with 70% ethanol. After centrifugation the pellet was air-dried and 200 µl DNA Hydration Solution (Qiagen) added, vortexed briefly and left overnight at room temperature to allow DNA to dissolve. DNA concentration was then measured on a NanoDrop2000 and frozen at -20 °C.

Primers for amplification of a HAP1 genomic DNA template, with a 537 bp product:

Forward primer: 5' GACGGAGGAGATCCAGTTCG 3'
Reverse primer: 5' TTCTTTAGCTGTGCAGTCGC 3'

Annealing temperatures were calculated using the NEB T_m calculator for the PCR reaction. For a total volume of 25 µl the following was prepared: 1.5 µl of 10 µM forward primer, 1.5 µl of 10 µM reverse primer, 5 µl Q5® polymerase buffer, 5 µl Q5® enhancer, 0.5 µl genomic DNA template, 0.5 µl 10mM dNTPs and 0.5 µl Q5® polymerase. The PCR thermocycler was set up as follows:

98 °C – 30 sec
98 °C – 10 sec
Annealing temperature as calculated – 30 sec] x35 cycles
72 °C – 30 sec
72 °C – 2 minutes
4 °C hold

PCR products were analysed on a 1.8% agarose gel and the correct product was gel extracted, yielding ~20 ng/µl DNA template. The cleavage reaction comprised 100 ng of the gDNA template, 20 ng of sgRNA, 1 µl Cas9 buffer, 1 µl 10x BSA and 1.5 µl Cas9 (Clontech) made up to 10 µl in nuclease free water. A positive control used 2.5 µl of the 2 kb control fragment (40 ng/µl) and 1 µl control sgRNA (20 ng/µl) that was run simultaneously to ensure nuclease efficacy. The test samples and control were incubated for 37 °C for 1 hour followed by 70 °C for 10 minutes. 5 µl were then visualised on a 1% TBE agarose gel to determine whether the expected banding pattern from successful cleavage was obtained. sgRNA-guided Cas9 cleavage of the DNA template would result in three bands – the largest for the uncut ~ 500 bp template and two lower bands denoting the smaller cleavage products and showing <100% cutting efficiency. sgRNA that did not successfully mediate cleavage would show as a single band of ~500 bp.

CRISPR/Cas9 knockout in cell lines using designed sgRNA

The cloning backbone used was pSpCas9(BB)-2A-GFP (PX458), shown in Figure 2.7, which contains GFP and Cas9 genes. PX458 was transformed into D5-α *E. coli* and plated onto 100 µg/ml ampicillin containing agar and mini-preps used to obtain high yields of PX458. The resulting DNA was checked by agarose gel electrophoresis (0.5 % TBE) to demonstrate one clean product band at ~9,000 bp.

PX458 was then digested with BbsI, which cut at two locations to remove the small length of sequence between the U6 promoter and sgRNA scaffold (Figure 2.7B) and linearise the plasmid. 10 units of BbsI (10,000U/ml) and 1 µg PX458 were used along with NEBuffer 2.1. The reaction was incubated at 37 °C for 2 hours and inactivated for 20 minutes at 65 °C. Restriction was confirmed by gel electrophoresis and the product purified using the QIAquick PCR Purification Kit (Qiagen), with a small volume run on 0.5% TBE agarose gel to ensure the correct band size and estimate concentration (using the MassRuler Express Forward DNA ladder).

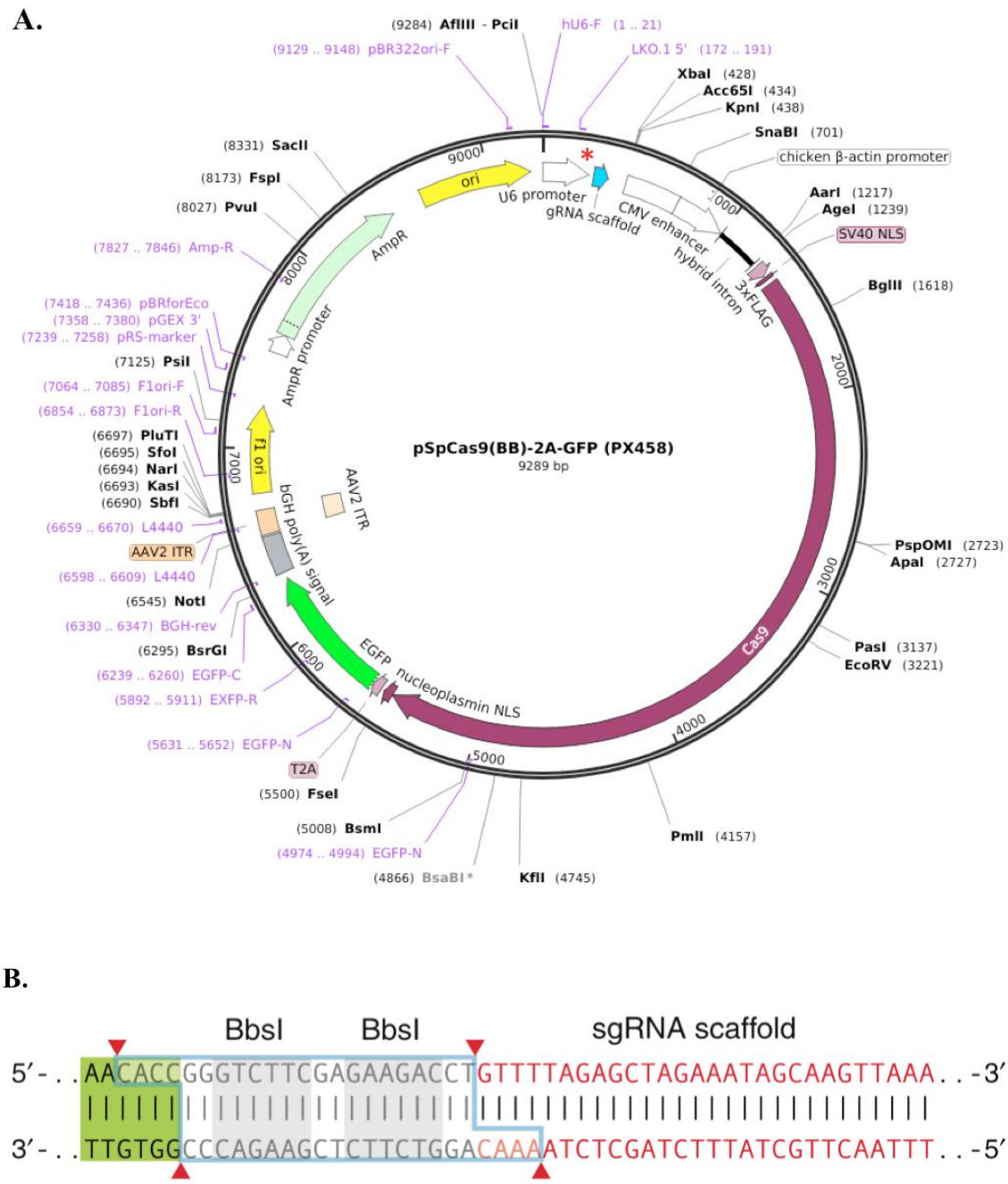


Figure 2.7 pSpCas9(BB)-2A-GFP (PX458). A. Vector map for PX458. Cloning vector for sgRNA template transformation into HAP1 cells (Addgene #48138). The red asterisk denotes the postion of sgRNA insertion after double BbsI digestion (detail in B.) The plasmid was created by Zhang and co-workers (Ran *et al*., 2013) and the plasmid map is available from https://www.addgene.org/48138/. **B. Restriction sites for BbsI digestion of PX458.** BbsI removes a short sequence of DNA shown by the blue box through recognition of two six base pair sequences, in grey (Ran *et al*., 2013). The annealed sgRNA oligomers (Figure 2.8) were then ligated into the vector.

To clone the sgRNA into PX458, two single stranded oligomers of the sgRNA (and reverse complement) with attached overhangs were designed and obtained from Sigma, using the template shown in Figure 2.8. The two oligomers were annealed by incubation at 95 °C for 5 minutes and then the temperature was ramped down to 25 °C with a decrease of 1 °C every 30 seconds, to form dsDNA with 4 bp overhangs at the 5' ends. The annealed oligomers were then ligated into digested PX458 using T4 DNA ligase. The oligomer mix was diluted 1:200 and two different volumes were used to obtain the optimal ratio insert:plasmid. 100 ng of BbsI restricted PX458 was added to 2 or 4 µl of diluted oligomer duplex, with the addition of 2 µl T4 DNA liagse buffer, made up to 20 µl with RNAse free dH_2O. The ligation mixtures were incubated at 37 °C for 2 hours and inactivated at 65 °C for 10 minutes.

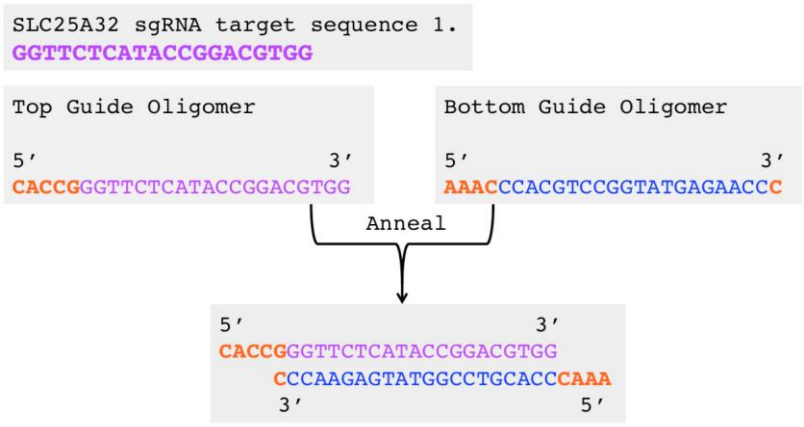


Figure 2.8 Template for sgRNA oligomers for cloning into PX458. The top oligomer comprises the sgRNA sequence (pink - example is SLC25A32 target sgRNA 1.) and a 5'extension of 5 bp (orange), whereas the bottom oligomer has a 4 bp 5' addition (orange) ahead of the reverse complement of the sgRNA sequence (blue), with a single base extension on the 3' end. These additions are required to allow ligation into the BbsI restricted PX458 at the postiton shown in Figure 2.7B.

The ligated PX458 constructs were then transformed into D5-α *E. coli,* plated onto 100 µg/ml ampicillin-containing LB agar and incubated at 37 °C overnight. 5 ml LB liquid cultures were then set up from all plates that yielded colonies, to be used for mini-prep processing. Plasmids were purified from the cultures using the Qiagen mini-prep kit and the products visualised on 0.5% agarose to estimate molecular weight. A selection of possible positive samples were chosen for DNA sequencing using the U6 forward primer. Sequencing was carried out by Genewiz UK to ensure correct insertion of the sgRNA into the PX458 plasmid.

Transfection of HAP1 cells with Cas9/sgRNA construct

After sequence validation of the construct the sgRNA containing PX458 plasmid was introduced into HAP1 cells by chemical transfection. 75,000 cells/well were plated in six-well plates ahead of transfection and incubated at 37°C, 5% CO_2 overnight in IMDM + 20% FBS. The transfection agent Lipofectamine™ 3000 was used and 3 wells were prepared for each construct, with a control well. For each well a lipofectamine mix was made: 125 µl OptiMEM and 7.5 µl OptiMEM, along with a DNA mix: 125 µl OptiMEM, 6 µl P3000 reagent and 3 µg PX458 construct. These two solutions were then mixed and incubated for 15 minutes at room temperature. The positive control (empty PX458) was used to validate efficient transfection of HAP1s using Lipofectamine™ 3000. Transfection mix was then added drop-wise to each well of DPBS washed HAP1s and incubated for 30 minutes at 37°C, 5% CO_2 before an additional 250 µl of OptiMEM was added. The cells were then incubated with the transfection reagents for 5 hours at 37°C, 5% CO_2 to facilitate PX458 construct uptake, then 5 ml IMDM + 20% FBS was added to provide growth conditions before incubation overnight. The following day the media was changed to remove the residual transfection reagent and 5 ml IMDM + 10% FBS added. The efficiency of transfection was observed by fluorescence from GFP expression on the plasmid.

The transfected cells were then fluorescence-activated cell sorted (FACS), to segregate the small proportion of successfully transfected, fluorescing cells from the background of HAP1s. A negative control was used, HAP1 parental cells at $1x10^6$ /ml, to prime the FACS equipment to sort the samples of transfected cells. Then each HAP1 line transfected with a PX458 construct was sorted and fluorescence data visualised by a dot-plot, showing signal scatter intensity and fluorescence intensity. Gating was used to select the fluorescing subpopulation, which were then seeded into pre-prepared 96 wells plates (IMDM + 20% FBS) at a density of ~1 cell/well to allow clonal expansion. Sorted cells were then incubated at 37°C, 5% CO_2 and checked for visible colonies. After ~7 days some colonies were observed, these were monitored until large enough to move into 24 wells plates, about 14 days after FACS. After a further 5 days these colonies were transferred to flasks (T25 or T75 as appropriate) with IMDM + 20% FBS. Once sufficiently expanded, and further increases in flask volume achieved, these colonies were banked for later testing by PCR, to validate the knockout.

PCR screen of transfected HAP1 cells

The commercially available SLC25A32 HAP1 knockout cell line (Horizon, HZGHC001715c10, SLC25A32 KO cells) was first raised in IMDM + 20% FBS, to encourage proliferation. Once established, along with the HAP1 parental line, these cells were conditioned in riboflavin-free DMEM + 10% FBS for one month. Morphology and growth profiles were monitored using Incucyte imaging software, demonstrating the differences between the SLC25A32 knockout and the parental HAP1. The SLC25A32 KO cells were analysed to check the SLC25A32 deletion by CRISPR. This was achieved

by PCR amplification of the gene region (exon 2) around the knockout and sequencing rather than by Western Blot; as the knockout was still likely to produce protein with a frame-shift (possibly by exon-skipping) that may be detected.

PCR primers
Forward primer: 5' AACCAATTGTACAGTGTCTG 3'
Reverse primer: 5' ATCATTCCACCTCATCTCTT 3'

Sequencing primer
Forward primer: 5' GTGATACAGAGCTAGAACTT 3'

For the CRISPR SLC25A32 knockouts I generated in HAP1 (described above) the same approach was used to screen the GFP positive clones, by amplifying the targeted region of exon 1 and sequencing it to determine whether any deletions had been achieved.

PCR primers
Forward primer: 5' GGATAAGAGTCCTCTCGTTGG 3'
Reverse primer: 5' CAGACTTCTCTACATACATGC 3'

Sequencing primer
Forward primer: 5' AAGAGTCCTCTCGTTGGTCC 3'

2.4. Mitochondrial preparation from human cell lines

2.4.1. Crude mitochondrial preparation – digitonised cells

Initial preparation of mitochondria from cell lines was achieved using digitonin solubilisation (Pello *et al*., 2008). ~5 x10^6 cultured cells were harvested and washed twice with phosphate buffered saline solution (PBS), all further steps were carried out at 4 °C. The pellet was resuspended in 200 µl cold PBS and 200 µl 8mg/mg digitonin added with incubation on ice for 10 minutes. 1 ml cold PBS was then added and the sample centrifuged at 10,000 xg for 10 minutes at 4 °C, the supernatant was then removed. Washing was repeated and the pellet resuspended in 100 µl 1.5 M aminocaproic acid, 50 mM Bis-Tris.HCl pH 7. n-dodecyl-β-D-maltoside (DDM) was added to a final concentration of 1% for analysis of individual respiratory complexes. After incubation on ice for 5 minutes the sample was centrifuged at 18,000 xg for 30 minutes at 4 °C and the supernatant retained. It was mixed 1:10 with sample buffer (750 mM aminocaproic acid, 50 mM Bis-Tris.HCl pH 7, 0.5 mM EDTA and 5% Serva Blue G250). Samples were then aliquoted and stored at -20 °C.

2.4.2. Mitochondrial purification and fractionation from cultured cells

Mitochondria were purified and fractionated from cell lines according to the protocol described by Hirst and co-workers, developed from the Holt and co-workers technique (Bridges *et al.,* 2017; Minczuk *et al.,* 2011). Cultured mammalian cells were grown in multilayer flasks until confluent (~3 days) and harvested after DPBS washes with 50 ml TryplE Express (Gibco) per flask. DPBS containing 10% FBS was then used to neutralise the detachment reagent and cells were pelleted by centrifugation at 500 xg for 3 minutes. Flasks were washed with DPBS to obtain all cells and used to resuspend pelleted cells before another centrifugation step. Pelleted cells were then washed twice with cold DPBS and the final pellet kept on ice. The wet cell mass was measured and, prior to cell breakage, resuspended in 3 ml hypotonic buffer (20 mM HEPES, pH 7.8, 5 mM KCl, 1.5 mM $MgCl_2$ and 1mg/ml bovine serum albumin, BSA), plus protease inhibitor, per 1g of cells and incubated for 10 minutes on ice.

Hypotonic buffer was used to swell the cells to increase breakage efficiency and cells were disrupted using an Isobiotec cell homogeniser. The homogeniser was chilled and kept on ice throughout breakage. It was first washed with hypotonic buffer and cells were then disrupted by seven passes through the chamber, with a gap set to be smaller than the cell diameter (12 μm for 143B and 10 μm for HAP1 cells). Broken cells were kept on ice and 0.7 ml of 2.5X MSH buffer (20 mM HEPES, pH 7.8, 525 mM mannitol, 175 mM sucrose, 5 mM EDTA), plus protease inhibitor, added per 1 ml of broken cells. The homogeniser was then washed with 2 ml 1X MSH and steps repeated until the cell suspension was homogenised, before making the volume of broken cells up to 30 ml with 1x MSH.
To remove cell debris and nuclei, broken cells were centrifuged at 750 xg for 10 minutes at 4 °C. The supernatant was then re-spun if separation of the globular pellet was incomplete. The supernatant was then centrifuged at 10,000 xg for 10 minutes at 4 °C to pellet the mitochondria. The supernatant was re-spun and both brown mitochondrial pellets pooled and resuspended in 1 ml 1x MSH + 50 U/ml benzonase, for DNA digestion. The sample was then incubated on ice for 10 minutes. To further purify the mitochondria and remove contaminating cellular components, particularly endoplasmic reticulum (ER), a sucrose gradient was used - comprising three layers made up in SW40 Ti tubes (9/16 x 3¾"). 4 ml of 1.5 M, 4 ml of 1.0 M and 3 ml of 0.5 M sucrose (with 10 mM HEPES, 5 mM EDTA, pH 7.8) were assembled and mitochondria were then layered on top, any remaining volume being filled by 1X MSH. The gradient was then centrifuged at 185,000 xg, 4 °C, for an hour to separate the components within the mitochondrial sample by density. The mitochondria formed a brown band at the interface between 1.5 and 1.0 M sucrose and any contaminating ER was found at the 0.5-1.0 M interface. The mitochondrial band was then removed and resuspended in 4 volumes of 1x MSH.

If the sample was to be retained and used as mitochondria the protein concentration was measured using the bicinchoninic acid assay (BCA). Mitochondria were then centrifuged for 10 minutes at 10,000 xg, 4

°C, and the pellet resuspended in solubilisation buffer (1.5 M aminocaproic acid, 50 mM Bis-Tris, pH 7) to adjust the concentration to 3-5 mg/ml. Aliquots were then snap frozen and stored at -80 °C.

For mitochondrial fractionation, the mitochondrial pellet obtained after centrifugation was further processed immediately or stored at -80 °C. Mitochondria were resuspended in 2-3 ml sonication buffer (20 mM Tris, 200 mM NaCl, 10% ethylene glycol, pH 7.5). The sample was sonciated on ice using a microtip (amplitude of 25), in 10-second bursts, until 100 J was absorbed by the sample and it had clarified. The sonicated mix was then centrifuged at 444,000 xg at 4 °C (MLA 80) to separate the mitochondrial matrix and membranes.

2.5. Protein analysis techniques

2.5.1. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Cell lysate samples were separated by SDS-PAGE. Cells were lysed with 20 mM Tris.HCl (pH 7.5), 500 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% (v/v) Triton-X-100, 10% glycerol and 1.5 mM $MgCl_2$. Protease inhibitors (cOmplete™ inhibitor cocktail tablets) and benzonase (Millipore) were added and samples were incubated on ice for 30 minutes and then centrifuged at 10,000 xg for 15 minutes to pellet cell debris, with the supernatant retained. Protein concentrations were determined by BCA and the lysate then frozen at -20 °C in 1:1 SDS loading buffer (0.125 M Tris, 20% (v/v) glycerol, 4% (w/v) SDS, 0.005% (v/v) bromophenol blue, pH 6.8, with 10 mM dithioreitol (DTT).

Samples were loaded at 20-40 µg protein for SDS-PAGE, using a modified Laemmli protocol (Laemmli, 1970). NuPAGE™ 4-12% Bis-Tris Protein Gels were run at 200 V in 1X MES-SDS running buffer (ForMedium) for ~40 minutes, until the dye front reached the bottom of the gel. Kaleidoscope™ Precision Plus (BioRad) standard protein marker was loaded alongside to allow molecular weight estimation. Gels were either stained with Coomassie® Brilliant Blue for 15 minutes and destained to remove excess dye, or taken for wet transfer onto membranes for Western Blot analysis.

2.5.2. Blue Native PAGE (BN-PAGE)

BN-PAGE was undertaken using the protocol described by Schägger and co-workers (Schägger and von Jagow, 1991). Mitoplasts were prepared as described in 2.4.1. Mitochondria were solubilised in either 1.2% DDM or 8g/g digitonin and centrifuged at 10,000 xg to remove insoluble material. Samples were diluted in 2:1 in NativePage™ sample buffer + 0.03% Coomassie G-250 (Amnesco) before loading. Fractionated mitochondrial membranes were solubilised in 1.7% DDM and diluted in NativePage™ sample buffer in the same way. 10-well NativePAGE™ Novex® Bis-Tris (3-12%) were run at 80 V for 1 hour at 4 °C. The outer anode buffer comprised pre-chilled 1X NativePAGE™ Running Buffer and the inner cathode buffer contained pre-chilled 1X NativePAGE™ Running Buffer + blue 1X

NativePAGE™ Cathode Buffer Additive. Another cathode buffer was also prepared with a 0.1X NativePAGE™ Cathode Buffer Additive. The inner buffers were exchange after 1 hour and the run continued at 180 V until the dye front reached the end of the gel. The gel was then removed from the casing and washed in dH_2O to remove excess Coomassie® G-250, at either room temperature (or 4 °C for in-gel activity assays). If the gel was to be stained 50% methanol (v/v), 7% acetic acid (v/v) and 0.2% Coomassie® R250 (w/v) were added for 5 minutes and destained in 20% methanol and 7% acetic acid overnight.

2.5.3. Western Blotting

Once protein samples had been separated by gel electrophoresis they were transferred onto an Immobilon®-P polyvinylidene fluoride (PVDF) membrane using a wet transfer system (300 mA, 150 V) with transfer buffer (25 mM Tris, 190 mM glycine and 10% methanol, pH 8.3) for 60 minutes at 4 °C. After transfer membranes were blocked in 10 ml blocking buffer containing 5% (w/v) milk in PBS at room temperature for ~4 hours before primary antibody incubation overnight at 4 °C. Membranes may be further blocked overnight with primary antibody incubation for two hours at room temperature the following day. Primary antibodies were used diluted in blocking buffer + 0.02% (v/v) Tween-20. Membrane slices were incubated in 6 ml primary antibody dilutions, after which primary antibodies were removed and membranes washed thrice with PBS + 0.2% Tween-20 for 20 minutes at room temperature. Rabbit anti-RFK was obtained from Proteintech, FADS (rabbit) from Aviva, SLC25A32 (rabbit), NDUFV1 (rabbit) and SDHA (rabbit) from Abcam and all used at 1:1000.

For electrochemiluminescence (ECL) detection horseradish peroxidase (HRP)-coupled secondary antibodies were used and diluted in blocking buffer + 0.02% (v/v) Tween-20. Goat anti-rabbit (Promega) and sheep anti-mouse (GE Healthcare) were used at 1:3,000 in blocking buffer and membranes incubated for 1 hour at room temperature. Membranes were then washed thrice with PBS + 0.2% (v/v) Tween-20 for 20 minutes at room temperature and a finally with PBS. ECL Prime detection kit (GE Healthcare) reagents were used and chemiluminescence detected using X-ray film. Once repeated and optimised the signal intensities could then be quantified using fluorescence coupled secondary antibodies (see below). Loading controls were also used (Rabbit anti-VDAC1 from Abcam and mouse anti-DRP1 from Abnova) to account for the total protein loaded.

2.5.4. Quantitative Western Blotting

Once protein samples had been separated by electrophoresis they were transferred onto an Immobilon®-FL PVDF membrane using a wet transfer system, as describe above. After transfer membranes were blocked in 10 ml protein-free PBS blocking buffer (Pierce) at room temperature for ~4 hours before primary antibody incubation overnight at 4 °C. Primary antibodies were used at the appropriate dilution, ascertained from Western blot detection by ECL, and made in blocking buffer + 0.02% Tween-20. After

overnight incubation primary antibodies were removed and membranes washed thrice with PBS + 0.2% Tween-20 for 20 minutes at room temperature.

For quantitative detection fluorophore-coupled secondary antibodies were used. These light sensitive antibodies (goat anti-rabbit IRDye® 680RD or goat anti-mouse IRDye® 800CW, Abcam) were diluted 1:5,000 in blocking buffer + 0.02% (v/v) Tween-20 + 0.02% (w/v) SDS and membranes were protected from light and incubated for one hour. SDS was added as PVDF membranes were used rather than nitrocellulose membrane and extra detergent is suggested by LiCOR. Membranes were then washed thrice with PBS + 0.2% Tween-20 for 20 minutes at room temperature, also in the dark. After a final wash with PBS membranes were imaged on a LiCOR Odyssey® CLx Imager. Signal intensities were quantified using the software to provide a comparison between protein levels in different samples. Loading controls were also used (VDAC1, DRP1) to adjust the signals obtained for the total protein loaded and background readings made to normalise signals to the base fluorescence of the membrane.

2.5.5. In-gel activity assays

To visually assess the activity of the respiratory complexes present in cell lines (crude mitoplasts or purified mitochondria), solubilised samples were separated by BN-PAGE and exogenous reagents added to detect activity of individual complexes, as described by Dabeni-Sala and co-wokers (Zerbetto *et al.*, 1997). For each gel 25 ml of reagent mix was added.

Complex I

Complex I activity was detected using the electron acceptor nitroblue tetrazolium (NBT), that forms a purple, in situ precipitate when reduced. BN-PAGE gels were washed in cold dH_2O to remove excess stain and retain activity. The reaction mixture of 0.5 mg/ml NBT and 100 μM reduced nicotinamide adenine dinucleotide (NADH) in 20 mM Tris pH 7.4 was added and after 30 minutes incubation at room temperature proteins with an active flavin cofactor - including complex I and the citric acid cycle oxoglutarate dehydrogenase complex (OGDH) – were highlighted by a purple band corresponding to their migration position. Once developed the gels were rinsed thoroughly and imaged.

Complex II

Complex II activity was assayed using 84 mM succinate, 2 mg/ml NBT, 10 mM sodium cyanide and 0.2 mM phenazine methosulphate (PMS) in 20 mM Tris.HCl, pH 7.5. Succinate acts as the electron donor to complex II, which reduces NBT, forming a visible purple band. SDHA subunit of complex II contains FAD, which re-oxidises succinate to fumarate, with electrons passed on to the exogenous reagent NBT.

Complex IV

The activity of complex IV was evaluated using 0.4 mg/ml 3-3 dibenzamidine (DAB) and 1 mg/ml cytochrome c in 50 mM phosphate, pH 7.2. Active complexes were observed as brown bands after incubation for ~1 hour at room temperature.

2.6. Respiratory phenotype techniques

2.6.1. Extracellular flux analysis of oxygen consumption rates

Oxygen consumption rates (OCRs) were measured using a Seahorse XF96 extracellular flux analyser (Agilent) at 37 °C. 11,500 143B or HAP1 cells were plated per well into Seahorse Bioscience XF96 plates and incubated for at least 12 hours at 37 °C in 5 % CO_2. The medium was then exchanged for assay buffer, either standard DMEM (4.5 g/L glucose, 1 mM pyruvate, 32 mM NaCl, 2mM GlutaMAX, 15mg/L phenol red and 20 mM HEPES, pH 7.4 at 37 °C) or riboflavin-free DMEM (formulation without riboflavin). Cells were then incubated for a further 30 minutes in a CO_2 free incubator. The basal OCR was first established and measured for at least four cycles (5 minutes mix, 3 minutes measure) before a stress test was undertaken. Oligomycin to a final concentration of 1 μM was injected into each well and measured for at least four cycles to inhibit complex V. Then 0.5 μM carbonyl cyanide p-trifluoromethoxyphenylhydrazone (FCCP) (0.25 μM for HAP1s) was injected with the same number of measurements to uncouple the mitochondria and measure maximum respiratory capacity. Lastly antimycin A and rotenone were both added to a final concentration of 1 μM to inhibit complex I and complex III to measure non-mitochondrial respiration for at least four cycles. Injection timings are shown in Figure 2.8, along with different components of oxygen consumption measured. After assay completion, to account for differences in seeding density, plates were washed with 200 μl/well PBS followed by cell lysis in 20 μl of 150 mM NaCl, 50 mM Tris (pH 7), 1 mM ethylenediamine tetra-acetic acid (EDTA), 1% Triton-X-100 per well, for 30 minutes. Protein concentrations were determined by the BCA assay. These values were then used to correct the OCR data (pmol/min) and extracellular acidification rate (ECAR, mpH/min) obtained.

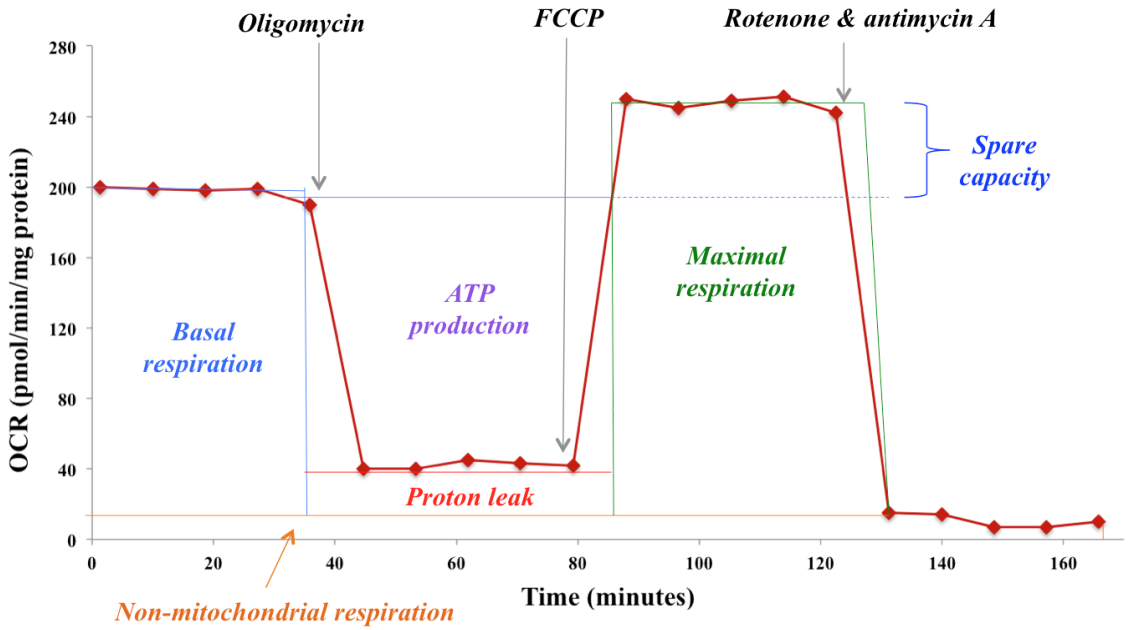


Figure 2.8 Example extracellular flux OCR data for a stress test with adherent cells. Injection points for oligomycin, FCCP and rotenone & antimycin A are shown on a typical OCR trace. The different components of oxygen consumption are labeled, showing how the stress test can be used to assess OCR: oligomycin inhibits ATP synthase, FCCP uncouples the mitochondria to allow maximal respiration and rotenone & antimycin A injection inhibits the ETC and highlights the residual OCR attributed to non-mitochondrial respiration.

2.6.2. Mitochondrial coupling, assessing membrane potential in live cells

HAP1 cells were plated into six-well plates the day before the experiment at 250,000/well. Then they were harvested by trypsinisation, pelleted by centrifugation (400 xg) and resuspended in 1 ml DPBS. 0.5 µl 50 nM tetramethylrhodamine (TMRM) and 20 µl 1 mg/ml Hoechst stain was added and samples incubated for 20 minutes at 37 °C, 5% (v/v) CO_2 to allow uptake of the dyes. Hoechst stain was used to stain the cell nucleus and correct the fluorescence measured for cell number. After incubation cells were pelleted to remove excess dye and washed twice with 1 ml DPBS. The final sample was then resuspended in ~200 µl DPBS.

Mitochondrial coupling was evaluated using a Chemometech NucleoCounter® NC-3000™, with 30 µl of cell solution loaded onto NC-Slide A2™ slides. Hoechst (excitation at 346 nm, emission at 397 nm) and TMRM (excitation at 488 nm, emission at 570 nm) uptake was then measured. The coupling of each sample was then assessed by comparison of the TMRM fluorescence to the control samples. The addition of 0.5 µl 2mM FCCP (10 minute incubation) uncoupled the mitochondria and by applying a gate to the data the proportion of uncoupled mitochondria could be assessed. This gate was then applied to all the data to determine the percentage of each population that was coupled and uncoupled.

2.7. Protein analysis by Mass Spectrometry

2.7.1. Stable-isotope labelling of amino acids in cell culture (SILAC)

Proteomic analysis by SILAC was undertaken as described by Walker and co-workers (Andrews *et al.*, 2013). ^{15}Nitrogen labelled lysine and arginine residues were incorporated into one sample to generate a shift in peptide fragment peaks, to allow a direct comparison to an unlabelled sample – generating a ratio of heavy to light for individual proteins. To achieve this, growth media lacking lysine and arginine was used – SILAC DMEM (Life Technologies). Heavy labelled amino acids were added as described below. Riboflavin free SILAC DMEM was custom synthesised by Life Technologies. Dialysed FBS was used as the growth supplement as it did not contain amino acids, to prevent unlabelled Arg and Lys being introduced into the heavy-labelled sample. Unlabelled proline was added at a final concentration of 200 μg/ml to all samples to prevent transamination of ^{15}N-Arg to ^{15}N-Pro during passages. Heavy-labelled amino acids were prepared in DPBS and ^{15}N-Arg added to the growth media at 0.398 mM and ^{15}N-Lys at 0.798 mM, to mirror concentrations found in control growth medium. Media was filtered sterilised after all additions.

The cell line under investigation and its appropriate control were grown in SILAC DMEM with either heavy labelled or unlabelled Arg and Lys, the experiment was then repeated with the opposite labelling for both a 'forward' and 'reverse' result. Therefore data was collected as the ratio of the heavy-labelled proteins within the cell line of interest compared to an unlabelled control, and vice versa. Cells were labelled for at least one week to ensure complete turnover of proteins and successful heavy labelling.

~$5x10^6$ cells were harvested and the pellets washed twice in DPBS. To determine the protein concentration, cells were resuspended in a small volume of DPBS and assessed by the BCA assay. A 1:1 mass ratio of heavy and unlabelled samples was then mixed prior to processing, and pelleted. In addition, a heavy-labelled only sample was processed to allow labelling efficiency to be checked.

Samples were then digitonised to form a crude mitochondrial sample as described in section 2.4.1. The protocol related to processing of 0.8 mg protein and therefore volumes were adjusted accordingly. Sample buffer was not added; instead samples were reduced and alkylated ahead of SDS-PAGE and analysis by mass spectrometry. An equal volume of 2x gel sample buffer (1x GSB: 20% (v/v) glycerol, 100 mM Tris pH 8, 2 mM EDTA, 2% (w/v) SDS) was added to the sample and tris-2-carboxyethyl-phosphine hydrochloride used at 5 mM final to reduce the sample, for 30 minutes at 37 ^{o}C. Iodoacetamide was added to 15 mM for alkylation of samples, incubated for 30 minutes at room temperature in the dark. DTT was used to quench excess iodoacetamide, to a final concentration of 30 mM.

Samples were then separated by SDS-PAGE and stained before bands were submitted for mass spectroscopy analyses. Sample lanes were divided into ~20 equal slices.

2.7.1.1. Sample preparation

Gel slices were further cut into four segments and rinsed twice in 100 µl HPLC-grade H_2O, then 100 µl 20 mM Tris.HCl (pH 8), both for one hour. Partial dehydration was achieved by 30-minute incubations in 100 µl 50% (v/v, Fluka) acetonitrile and 20 mM Tris.HCl (pH 8). Full dehydration was performed using 20 µl 100% (v/v) acetonitrile for 10 minutes followed by 30 minutes in a Savant Speedvac® vacuum concentrator at 37 °C. Gel pieces were digested with 10 µl of 12.5 µg/ml bovine trypsin in 20 mM Tris.HCl (pH 8) and 5 mM $CaCl_2$ with incubation overnight at 37 °C. Tryptic peptides were then extracted using 15 µl of 20 mM Tris.HCl (pH 7.4), 60% (v/v) acetonitrile and 4% (v/v) formic acid for one hour. The extracted solution was transferred to a new tube, and the gel was re-extracted for 1 hour. In the case of BN-PAGE gels (for complexomics, section 2.7.2.), the extracted solution was further vacuum dried to increase the concentration of peptides and remove excess acetonitrile, that would prevent peptides from adhering to the reverse-phase high-performance liquid chromatography (RP-HPLC) column. Finally, the samples were re-suspended in 15 µl of 2% (v/v) acetonitrile and 0.1% (v/v) formic acid.

2.7.1.2. Liquid chromatography MS/MS Orbitrap mass spectrometry

The peptide mixtures were fractionated by RP-HPLC on an Acclaim PepMap nanoViper C18 column (75 µm internal diameter, 100 mm length) in a Proxeon Easy-nLC system using a gradient of 4.9-39.2% acetonitrile in 0.1% (v/v) formic acid over 84 minutes, at a flow rate of 300 nl/min. The eluate was transferred in line to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific). Data were acquired from 400 to 16000 m/z for precursor ions and the ten most abundant precursor ions fragmented by higher-energy collisional dissociation in nitrogen. Analyses were performed by Dr Shujing Ding from the MRC MBU, University of Cambridge.

2.7.1.3. SILAC protein identification and quantification

Peptide mass and fragmentation data were processed with MaxQuant (version 1.5.0.12; Cox & Mann, 2008) and the integrated Mascot search engine (Matrix Science) used to identify proteins by comparison to the Swiss-Prot human database. Ratios of heavy to light-labelled proteins were calculated by MaxQuant; each protein ratio represents the median of ratios of all assigned peptides for that protein. MaxQuant ISO-MS/MS peptide ratios, calculated from intensities spanning regions in the m/z retention time plane corresponding to isotopic patterns, were used only when fewer than three other peptide ratios were obtained. Quantitative data were processed using Perseus software (version 1.3.0.4; Tyanova *et*

al., 2016) and proteins identified in a contaminant database or decoy database (created by MaxQuant by reversing protein entries) were excluded. Proteins lacking unique peptides were also removed. For protein experiments in which protein labelling was in the reverse orientation the reciprocal value was calculated. Base 2 logarithmic values are presented on scatter plots.

2.7.2. Complexomics

After the completion of BN-PAGE (section 2.5.2.) each sample lane was cut into 64 1mm thick equal sized bands and submitted for mass spectrometry for complexome analysis according to Brandt and co-workers (Heide *et al.*, 2012). All analyses were performed by Dr Mike Harbour from the MRC MBU, University of Cambridge.

Samples were digested by trypsin (section 2.7.1.1.), resolved using RP-HPLC on an Acclaim PepMap nanoViper C18 reverse-phase column (75 µM x 100 mm) and introduced onto a LTQ Orbitrap XL™ (Thermo Fisher Scientific) with a nanospray inlet interface, coupled to an in-line Proxeon EasynLC nano chromatograph. Buffer A (5% acetonitrile and 0.1% formic acid) and buffer B (95% acetonitrile and 0.1% formic acid) were used to set up a gradient up to 40% acetonitrile over 84 min with a flow rate of 300 nl/min. Slices were analysed in turn, with the highest molecular weight slice down to the lowest, without washes to maintain the same chromatographic conditions throughout. The ten highest scoring peptides (400-2000 m/z) were chosen for collision-based fragmentation. To improve the range of peptides analysed a dynamic exclusion list was generated to exclude previously analysed peptide fragments for a specified time.

Mass spectrometry data was outputted as a .RAW file (~120 Mb for each sample). Initially files were processed by Proteome Discoverer software (Thermo Fisher Scientific) and the in-house Mascot server was used to identify the sequences of Uniprot-database-derived peptides from their fragments. Additionally, trypsin was specified as the digesting agent and settings adjusted to allow one missed cleavage site (lysine and arginine). A theoretical peptide sequence is then created. To account for noise within experimental samples a flexibility of 10 ppm around the calculated peptide masses was allowed and a window set at 0.5 Da. Four types of peptide modifications were included: oxidation of methionine, the acrylamide adduct of cysteine (propionamide cysteine), acetylation of protein N-termini and formylation of N-termini.

The results from Mascot were then used by Proteome Discoverer for Peptide-Spectrum Match Validation - false discovery rates of 0.1 (strict) and 0.05 (relaxed) were used to derive the traffic light colour-coding scheme, which indicates the confidence of each Mascot identification. The remaining stages defined the complexomics approach, identifying and quantifying the peptide chromatographic peaks for relative quantification. In this way complexomics is used to quantify each peptide, of each

protein, within each slice – completed by Proteome Discoverer for each of the 64 slices/sample. Finally each of the 64 results was combined into a single MultiConsensus report. An algorithm was applied to convert relative peptide quantification into relative quantification of proteins: the three highest intensity peptides for each protein are selected, the mean value taken to generate a number representative of the whole protein. If less than three peptides were present the average of those present was taken. This was repeated for each slice to form a profile of each protein throughout the lane.

Protein profiles were then further analysed using the criteria from Brandt and co-workers (Heide *et al.*, 2012). The Pearson uncentered correlation coefficient (PUCC) for each possible pair of protein profiles was used to calculate single-linkage hierarchical clustering. Profiles were scaled and normalised as a proportion of the highest count recorded. For data analysis of protein complexes within the samples the final result was used to generate a heat map, for ease of visualisation, whereby the background colour of each cell (representing the protein and slice number) was coloured depending upon its value. A range was built around 0 for white, 0.7 for yellow and 1 for red, using a visual basic for applications (VBA) Macro to generate a heatmap showing the relative distribution of each protein identified across the 64 slices.

2.8. Gene expression techniques

2.8.1. Total cell RNA purification

RNA from HAP1 lines was purified using the RNeasy kit (Qiagen), starting from a 70% confluent T75 flask of cells. Once harvested cell number was ascertained for purification of 3-4 $x10^6$ cells/sample.

Following the kit instructions 40 mM DTT was added to an aliquot of RLT buffer before use. Cells were disrupted with 350 µl RLT buffer and vortexed before homogenisation using Qiashredder spin columns. Centrifuging for 2 minutes at 8,000 xg was sufficient to process the sample; the same volume of 70% ethanol was then added to the lysate and mixed to remove contaminating DNA. The sample was then passed through a RNeasy spin column by 15-second centrifugation at 8,000 xg and flow-through discarded. 700 µl RW1 buffer was then added to wash the column, followed by 500 µl RPE buffer, and another 500 µl with a 2-minute centrifugation. To ensure the membrane was completely dry it was re-spun for 1 minute without additions. RNA was then eluted with 30 µl RNAse free dH_2O, aliquoted and snap frozen before being stored at -80 °C. For each HAP1 and SLC25A32 KO lines three biological repeats were purified, with a HAP1 and SLC25A32 KO pair purified side-by-side.

2.8.2. RNA quality analysis

RNA quality was then assessed by the Agilent high sensitivity RNA ScreenTape system according to the manufacturer's guide. High sensitivity RNA ladder and two-fold diluted RNA samples were prepared in high sensitivity sample buffer at 2:1 ratio. Samples were vortexed for one minute and centrifuged before heating to 72 ℃ for three minutes, incubated on ice for two minutes and centrifuged to collect the sample. Samples were then loaded into a 2200 TapeStation instrument and analysed for RNA quality, with a requirement for greater than a 7/10 quality score to proceed with transcriptomics analysis.

2.8.3. Transcriptome profiling and analysis

500 ng of each RNA sample in 25 µl was submitted to Eurofins for Favourite Transcriptome Profiling to provide three repeats for each cell line. RNA quality was assessed on receipt by Eurofins and the RNA prepared using the TruSeq stranded mRNA library preparation kit. Sequencing was then undertaken on an Illumina HighSeq 2500 with 100 bp pair ended reads (sequenced in both directions). Processing of transcriptome data was performed in house by Dr Alan Robinson from the MRC MBU, University of Cambridge, using DESeq2 (Love *et al.,* 2014) to obtain parameters for fold change and adjusted P-value for analysis of changes in transcription profile.

2.9. Preparations of complex I containing samples from *Yarrowia lipolytica*

2.9.1. Generation of NDUFS2 mutants in *Y. lipolytica*

The mutant strains of NDUFS2 (NUCM) in *Y. lipolytica,* shown in Table 2.7, were designed and generated by Dr Andrew Jones using the Δ*nucm* strain produced by Brandt and co-workers (Grgic *et al*., 2004). This strain was produced from the parental GB10 strain, developed with a His-tagged NDUFS3 subunit and internally directed alterative NADH dehydrogenase (NDH2) to allow for compensation of complex I mutants (Kashani-Poor *et al*., 2001; Kerscher *et al*., 2001). Point mutations were made by PCR to the sequence of NUCM cloned into pUB26, to be expressed under the control of its native promoter. This plasmid includes ampicillin resistance for selection in bacteria and hygromycin for selection in *Y . lipolytica*. After amplification of the mutant constructs in *E. coli* and sequencing of the plasmids, the plasmids were transformed into Δ*nucm* (*ΔNDUFS2*) *Y. lipolytica* by electroporation. Expression of complex I and activity was assessed by BN-PAGE of purified mitochondria compared to the wild type GB10 strain.

Residue number (*Y. lipolytica*)	Codon sequence	Amino acid	Mutated codon sequence	Resulting amino acid	Human residue identity
91	CAT	H	ATG	M	H55
95	CAC	H	ATG	M	H59
121	CGA	R	ATG	M	R85
141	CGT	R	ATG	M	R105
144	TAC	Y	TTC	F	Y108
192	TCA	S	GTA	V	T156
196	GAT	D	AAT	N	D160
226	CAC	H	ATG	M	H190

Table 2.7 NUCM *Y. lipolytica* mutants and equivalent residues in the mature human subunit NDUFS2. Nine mutants were designed in NUCM and expressed in Δ*nucm Y .lipolytica* GB10 to assess complex I activity. Mutations were chosen in the putative Q-binding channel and represent a number of residues with evidence of ligand binding or iron sulphur cluster interactions and some with location and properties that suggest a role in ubiquinone binding.

2.9.2. Mitochondrial membrane preparation of *Y. lipolytica* NDUFS2 mutants

Y. lipolytica mutants were grown in aliquots of 25 ml 2x yeast extract peptone dextrose (YPD) medium (20g/L yeast extract, 40g/L peptone, 40g/L glucose, pH 5.5) containing 100 μg/ml hygromycin. Cultures were incubated at 30 °C, shaking at 225 rpm, for about 72 hours, until OD_{600} reached 10-20, then used to inoculate 2 litre flasks containing 500 ml 2x YPD + 50 μg/ml hygromycin to OD_{600} of 0.3. The flasks were incubated at 30 °C, shaking at 225 rpm, for about 96 hours, until OD_{600} reached 30. *Y. lipolytica* cells were harvested by rounds of centrifugation at 7,000 xg for 15 minutes at 4°C and washed twice with cold dH_2O. Pellets were stored at -80 °C until membrane processing.

Mitochondrial membranes were prepared as described by Varghese *et al*., but used a different cell disruptor due to smaller sample volumes (Varghese *et al*., 2015). A frozen pellet was defrosted and resuspended in twice the volume (relative to its mass) of breaking buffer (20 mM MOPS, 400 mM Sorbitol, 5 mM EDTA, 2 mM benzamidine, 2% (w/v) BSA, 5 mM amino-caproic acid, pH 7.2) + 2 ml 200mM PMSF. Cells were then disrupted in a Z-plus 2.2 KW cell homogeniser (Constant Cell Disruption Systems) at 35 Kpsi, for three rounds, to ensure maximum breakage. Cell debris was then removed by centrifugation (10,100 xg, 30 minutes, 4 °C) and the supernatant retained. Mitochondrial membranes were then pelleted by high speed centrifugation at 200,000 xg for 1 hour at 4 °C. Pellets were combined, homogenised and washed in wash buffer (20 mM MOPS, 400 mM Sorbitol, 2 mM benzamidine, 5 mM amino-caproic acid, pH 7.2) twice at 200,000 xg, 45 minutes, 4 °C before finally being resuspended in 20 mM NaH_2PO_4, 150 mM NaCl and snap frozen in 25 ml aliquots.

2.9.3. Complex I purification from *Y. lipolytica* mitochondrial membranes

Complex I was isolated from *Y. lipolytica* as described below, based on the protocol used by Varghese *et al*. that was developed from the protocol of Brandt and co-workers (Varghese *et al.,* 2015; Kashini-Poor *et al.,* 2001). Modifications include addition of 0.02% Asolectin/3-[(3-cholamidopropyl)-dimethylamminio]-1-propane sulphonate (CHAPS) to Buffers A and B and a Superose™ 6 Increase gel filtration column was used. *Y. lipolytica* membrane aliquots were thawed (500-800 mg protein) at 4 °C and centrifuged at 235,000 xg for 30 minutes at 4 °C to pellet the membranes. These pellets were then homogenised in Buffer γ to 20 mg/ml (Table 2.8) and protease inhibitors (2x cOmplete™ protease inhibitor cocktail tablets) were added. The mixture was stirred with 2.9% DDM (from a 10% stock) added drop-wise and solubilised for 30 minutes on ice, before centrifugation at 235,000 xg for 45 minutes at 4 °C. The supernatant was retained and solid NaCl was added to a final concentration of 400 mM and imidazole from a neutralised 1M stock was added to 20 mM. The solubilised membranes were then filter sterilised and loaded onto a Nickel-Sepharose column (GE Healthcare) to bind the His-tagged complex I.

	Buffers			
Components (mM)	**γ**	**A**	**B**	**Gel Filtration**
Na-MOPS				20
Na_2PO_4	20	20	20	
NaCl	50	400	400	150
Imidazole		52	140	
Asolectin/CHAPS (w/v)		0.02%	0.02%	
DDM (w/v)		0.1%	0.1%	0.05%
Glycerol (v/v)				10%
pH at room temperature	7.45	7.20	7.20	7.45

Table 2.8 Buffer composition for isolation of complex I from *Y. lipolytica*. Purification required four buffers, all kept at 4 °C. Asolectin/CHAPS and DDM additions were made on the day with buffers A, B and gel filtration filtered before use.

The column was then washed and bound protein eluted using an ÄKTA Prime liquid chromatography system (GE Healthcare). The column was washed at 2 ml/min with Buffer A (Table 2.8) until the conductance reading stabilised. Buffer B (Table 2.8) was then used to elute the bound His-tagged complex I. 3 ml fractions with the highest protein absorbance from the eluted volume were pooled (~10 fractions) and concentrated (Millipore 50K MWCO, 50 ml) at 5,800 xg until 2x 500 µl concentrated aliquots were obtained. The sample was then injected onto a Superose™ 6 Increase gel filtration column (GE Healthcare) connected to an ÄKTA Explorer fast protein liquid chromatography system (GE Healthcare). 0.5 ml fractions of protein were collected by elution with gel filtration buffer (Table 2.8) at a rate of 0.5 ml/min. Fractions from the second peak in absorbance at 280 nm were pooled (the first peak represented aggregated protein). Peak fractions from multiple 500 µl injections were then

concentrated together to ~250 µl. Isolated complex I was aliquoted into 10 µl samples and snap frozen before storage at -80 °C.

2.9.4. Preparation of alternative oxidase (AOX) for use in proteoliposomes

AOX was purified from *E. coli* (FN102/pTbAO strain) mitochondrial membranes at 4 °C as described by Hirst and co-workers (Jones *et al*., 2016). 20 ml of membranes were solubilised at 6 mg/ml protein in 25 mM Tris.HCl (pH 8.0), 200 mM $MgSO_4$, 1.4% (w/v) *n*-octyl-glucopyranoside (Anagrade, Anatrace) and 20% (v/v) glycerol for 1 hour then centrifuged at 165,000 ×g for 30 minutes. The supernatant was loaded onto a ~8-mL column of Strep-Tactin® Superflow® high capacity resin (IBA GmbH) pre-equilibrated in strep buffer (20 mM Tris.HCl, 50 mM $MgSO_4$, 160 mM NaCl and 20% (v/v) glycerol, pH 8.0 at 4 °C). AOX was eluted in strep buffer supplemented with 2.5 mM desthiobiotin (Sigma-Aldrich) and 0.042% DDM. Pooled AOX-containing fractions were concentrated 10-fold, then dialysed for six hours against 2 L of strep buffer plus 0.042% DDM.

2.9.5. Preparation of proteoliposomes from *Y. lipolytica* complex I

Proteoliposomes were generated using the protocol described by Jones *et al.,* developed from that of Roberts and Hirst (Jones *et al*., 2016; Roberts & Hirst, 2012). Proteoliposomes were prepared in 50 mM KCl, 10 mM Tris.SO_4 at pH 7.5 (PL buffer) at 4 °C. Stock phospholipids of 25 mg/ml (bovine heart, Avanti Polar Lipids) in chloroform were mixed in a glass homogeniser: phosphotidylcholine (PC), phosphotidylethanolamine (PE) and cardiolipin (CL) at 8:1:1 for a total of 10 mg phospholipid. Q_{10}, also in chloroform, was also added at 0, 12.5, 25, 50, 75, 100 or 150 nmoles (equivalent to 0-15 mM in lipid). Chloroform was removed by evaporation under $N_{2(g)}$ and further evaporated by applying a vacuum. The mixture was then resuspended in ~675 µl PL buffer (depending on protein volumes added in later steps for a final volume of 1 ml) and extruded nine times through a 0.1 µM pore tracked edge membrane (Whatman). All samples were then chilled to 4 °C overnight prior to the next steps.

Lipids were partially solubilised by 160 µl octyl-glucoside (OG, aqueous 10% stock), sonicated for 10 minutes and chilled for 10 minutes on ice. All following steps were done at 4 °C. 0.2 mg AOX was added along with 0.2 mg of isolated *Y. lipolytica* complex I. After 10 minutes incubation ~100 µl of washed biobeads (Bio-Rad SM2 Biobeads) were added to each sample and agitated for 4 hours - ~100 µl of fresh beads were added every hour. Biobeads were then removed using a micro bio-spin column (Bio-Rad) and the suspension centrifuged at 57,500 xg for 45 minutes to collect the proteoliposomes. The supernatant was removed and the pellet resuspended in 100 µl cold PL buffer. The final volume was recorded (for component quantification), aliquoted and snap frozen for storage at -80 °C. This protocol typically yielded ~200 µl proteoliposomes at 1.5 mg/ml total protein.

2.10. Kinetic measurements

In general, a master mix containing the assay buffer, protein and electron acceptor was prepared and 150 µl pipetted into a well of a 96-well plate. The reaction was started by the addition of 50 µl NADH (usually 200 µM final concentration) into each well, after a two-minute incubation at 32 °C. Kinetic measurements were made using a Spectramax 384 plate reader (Molecular Devices) at 32 °C and all data were processed using SoftMax Pro (Molecular Devices) to obtain activity rates, expressed as µmol reagent oxidised/min/mg protein.

2.10.1. Kinetic measurements on *Y. lipolytica* mitochondrial membranes

Assay buffer comprised 10 mM TrisHCl.SO_4, 50 mM KCl, pH 7.5 at 32 °C. Concentrations given below represent the final concentration in the assay, taking into account the dilution of reagents when the master mix and NADH solutions were mixed.

Complex I specific activities in mitochondrial membranes were measured with a range of electron acceptors and deamino NADH (dNADH) was added at 200 µM. The quinones used were decylubiquinone (DQ, 200 µM) and Q_1 at (100 µM). For NADH:O_2 assays a saturating concentration of cytochrome c was added (8 µM). NADH oxidation was monitored at 340-380 nm (ε = 4.81 mM^{-1} cm^{-1}) and for all assays *Y. lipolytica* membranes were used at 30 µg/ml. 2 µM piericidin A was added to inhibit quinone reduction at the Q-site.

Additionally the artificial electron acceptors 3-acetylpyridine adenine dinucleotide ($APAD^+$) and potassium hexacyanoferrate (III) (FeCN) were used. $APAD^+$ was added at 500 µM and 100 µM dNADH was used with the addition of 2 µM piericidin A to inhibit Q-site activity. The formation of APADH was followed at 400-450 nm (ε = 3.16 mM^{-1} cm^{-1}). For FeCN measurements FeCN was added at 1 mM, dNADH at 200 µM and piericidin A at 2 µM. NADH oxidation was monitored at 340-380 nm (ε = 4.81 mM^{-1} cm^{-1}).

2.10.2. Kinetic measurements on isolated *Y. lipolytica* complex I

All isolated complex I assays were carried out in 20 mM Tris.HCl, pH 7.5. Concentrations given below represent the final concentration in the assay.

NADH:DQ activity assays used complex I at 0.25 µg/ml, DQ at 200 µM and NADH at 200 µM. 0.075% (w/v) asolectin-CHAPS (from a 5% (w/v) stock) was also required in the complex I containing solution. NADH oxidation was assessed at 340-380 nm (ε = 4.81 mM^{-1} cm^{-1}) over several minutes.
Production of reactive oxygen species (ROS) was assessed using the probe Amplex Red (AxR, Invitrogen), which reacts with hydrogen peroxide. The master mix contained 0.075% Asolectin-

CHAPS, 10 μM AxR and 10 μg/ml complex I, with 10 U/ml horseradish peroxidase (HRP) to catalyse the reaction. The formation of the fluorescent product resorufin was detected by following A_{557} and A_{620} ($\varepsilon = 51.6\ mM^{-1}\ cm^{-1}$). Superoxide dismutase (SOD) was present in the master mix at 10 U/ml, to convert all superoxide to hydrogen peroxide. 30 μM NADH was used to initiate the assay, after the two-minute incubation. Catalase insensitive (5000 U/ml) and complex I independent background rates were subtracted to calculate the actual rates of ROS production by complex I. This reaction was very slow and therefore rates were calculated in nmol resorufin produced/min/mg complex I.

2.10.3. Kinetic assays on *Y. lipolytica* complex I containing proteoliposomes

All proteoliposome assays were carried out in 50 mM KCl, 10 mM Tris.SO_4, pH 7.5 at 32 °C. PLs were diluted 1:500 in all assays. Concentrations given below represent the final concentration in the assay.

NADH:Q_{10} assays were carried out with PLs containing different concentrations of Q_{10}. 200 μM NADH was added to initiate NADH oxidation, which was monitored at 340-380 nm ($\varepsilon = 4.81\ mM^{-1}\ cm^{-1}$) over several minutes. NADH:Q_1 assays used the same concentration of NADH but required PLs prepared without Q_{10}. Q_1 was then added from diluted ethanol stocks for concentrations of 0, 2, 5, 10, 20, 40, 60, 80 and 100 μM (a constant ethanolic volume was used). 2 μM piericidin A was added to inhibit Q-site activity.

To determine the IC_{50} for piericidin for each mutant and WT, in the presence of saturating Q_{10}, the proteoliposomes with the [Q_{10}] closest to 10 mM in the membrane (after quantification) were chosen. Piericidin A was diluted in ethanol for concentration of 0-2 μM and a constant volume was added to the assay. NADH was added to 200 μM and the activity of each mutant over the range of piericidin concentrations was assessed at 340-380 nm ($\varepsilon = 4.81\ mM^{-1}\ cm^{-1}$).

2.11. Analytical methods

2.11.1. Quinone concentration in Q_{10} proteoliposomes

Q_{10} quantification was carried out as described by Hirst and co-workers (Fedor *et al*., 2017). 90 μl of ethanol (HPLC grade, Chromasolv) was added to 10 μl of proteoliposomes solution, sonicated for 1 minute and centrifuged at 16,300 ×g for 10 minutes. A total of 150 μl of supernatant was injected onto a Nucleosil 100–5C18 (Hichrom) column, and run at 30 °C at 800 μl/min on an Agilent 1100 series HPLC in 70% (v/v) ethanol, 30% (v/v) methanol, 0.07% (v/v) $HClO_4$, and 50 mM $NaClO_4$ with a retention time of 9 minutes. Concentrations were determined by comparison with known standards.

2.11.2. Complex I concentration in proteoliposomes

Total protein contents were quantified for complex I only proteoliposomes as described by Hirst and co-workers (Fedor *et al*., 2017), using the amido black assay, which is insensitive to detergents and high phospholipid concentrations (Kaplan *et al*., 1985). 10 µl of proteoliposome sample was diluted in 2 ml dH_2O. 0.2 ml 10% (w/v) SDS, 0.3 ml 1 M Tris/1% (w/v) SDS, pH 7.5 and 0.6 ml 104% trichloroacetic acid (TCA) were added sequentially with vortexing between. Samples were then incubated for at least 3 minutes, transferred onto a 0.45 µm pore filter (Millipore, HAWP02400) and placed on a Millipore sampling manifold. The sample was then filtered under vacuum and washed with 2 ml 6% TCA. After all samples had been filtered staining was done with 200 ml 0.1% (w/v) amido black 10B dissolved in methanol/glacial acetic acid/deionised water (45/10/45, v/v) for 3 minutes. Three one-minute rounds of washing in destain solution - methanol/glacial acetic acid/deionised water (90/2/8 v/v) was used to remove excess stain. Filters were then washed in 200 ml dH_2O, blotted dry and the blue dot excised before being cut into six. Stain was eluted by addition of 0.7 ml 25 mM NaOH/0.05 mM EDTA/50% (v/v) ethanol and incubation for one hour with vortexing. 200 µl of eluate was then read in duplicate at 630 nm.

The NADH:$APAD^+$ oxidoreduction assay was used to calculate the complex I content in all samples, as described by Hirst and co-workers (Jones *et al*., 2016). A master mix containing diluted PLs and 2 µM piericidin A was prepared, along with another solution of 100 µM NADH and 500 µM $APAD^+$. Solutions were incubated at 32 °C for two minutes and then mixed. APADH formation was followed by changes in A_{400} and A_{450} ($\varepsilon = 3.16\ mM^{-1}\ cm^{-1}$).

The complex I only standards were assayed with and without addition of 15 µg/ml alamethicin, which permeabilised the membrane to allow access to all complex I (orientated to the outside and inside of the PL). The percentage of complex I facing out of the PL could then be determined by comparison of the ± alamethicin NADH:$APAD^+$ activities. Turnover numbers (k_{cat}) for NADH:$APAD^+$ were calculated for each variant using the alamethicin-treated standard, complex I only, activity and the protein concentration from the Amido Black assay for these PLs (calculation described in section 5.5.1). The NADH:$APAD^+$ activities for batches of PLs containing different concentrations of Q_{10} were measured in the absence of alamethicin. The k_{cat} calculated using the standard complex I only PL, for each variant and WT, was then used with the corresponding NADH:$APAD^+$ activities from the batches of PLs (with different [Q_{10}]) to determine the concentration of outward facing, substrate accessible, complex I in each variant PL. Activities with Q_{10} and Q_1 could then be converted into specific activities.

2.11.3. Phospholipid concentration in proteoliposomes

Total phospholipid contents were determined using the Ames method (Ames, 1966), applied to the proteoliposome system as described by Hirst and co-workers (Jones *et al*., 2016). 100 µl of each sample

(at several dilutions up to 0.4 mg/phospholipid ml) or standard (0 to 500 μM KH_2PO_4) was transferred into a boiling tube along with 50 μl of methanol and 30 μl of 390 mM $Mg(NO_3)_2$ in ethanol. The tubes were heated over a roaring blue flame in a fume-hood until no further brown fumes were formed and the residue in the tubes had turned white (without boiling). The tubes cooled for 5 minutes, then 0.3 ml of 500 mM HCl was added and the tubes were lightly stoppered with glass marbles and incubated at 99 °C for 15 min. Once cooled, 0.7 ml of an aqueous solution of 114 mM ascorbic acid, 2.72 mM $(NH_4)_6Mo_7O_{24}$ and 400 mM H_2SO_4 was added; tubes were lightly stoppered and incubated at 37 °C for 1 hour. 200 μl aliquots were transferred into the wells of a quartz 96-well plate, the absorbance (from reduced phosphomolybdate) measured at 820 nm, and the phospholipid contents of the samples determined by reference to the standards, accounting for the fact that each PC or PE contributes one orthophosphate, whereas each CL contributes two.

Chapter 3

Investigating incorporation of the flavin mononucleotide cofactor into complex I

3.1. Introduction

3.1.1. Assembly of complex I

The assembly pathway for complex I has long been studied; however, due to its large structure and number of subunits elucidation of this intricate pathway has been gradual. Additionally, the enzyme is under dual genomic control; requiring the coordination of subunits encoded by both the mitochondrial and nuclear genomes (Vartak *et al*., 2014). The first organism used to study the pathway was the fungus *Neurospora crassa* (containing 43 different subunits – Marques *et al*., 2005), and those studies suggested that the membrane and matrix arm subunits form as separate subassemblies before joining to make the mature complex. Weiss and co-workers treated *N. crassa* with chloramphenicol to inhibit mitochondrial translation and observed accumulation of nuclear encoded subunits in a smaller complex; their identities were determined by Western Blotting and peptide mapping - this subcomplex also showed NADH oxidation activity (Freidrich *et al*., 1989). In the same way the membrane arm was shown to assemble in *N. crassa* grown in manganese-limited conditions (which disrupts mitochondrial function); EPR investigations demonstrated this subcomplex contained the terminal iron-sulphur cluster N2 but lacked other redox cofactors (Schmidt *et al*., 1992). Additionally, manipulation of *N. crassa*, through knocking out the NDUFS8 homologue, identified the first chaperone-like proteins (CIA30 and CIA84), found associated with the membrane domain intermediate but absent from the complete enzyme (Küffner *et al*., 1998). These studies highlighted not only the modular steps in complex I assembly but also the involvement of other proteins, giving rise to research into related assembly factors.

More recent work with mammalian mitochondria supported the modular concept and a pathway with the membrane and matrix arms joining together to form the mature enzyme formed the basis of more detailed proposals. Figure 3.1 shows the first assembly pathway, which begins with the formation of a 400 kDa subcomplex built around NDUFS2 and NDUFS3 and membrane-bound due to ND1. Other subunits are added, along with assembly factors NDUFAF3 and NDUFAF4. Separately, another subcomplex comprising membrane arm subunits is assembled with assembly factors NDUFAF1, ACAD9 and ECSIT to form a 460 kDa subcomplex. These two intermediates then combine, and with the addition of the more distal membrane domain subunits, ND4 and ND5, an 830 kDa subassembly is formed. This represents the full enzyme lacking the NADH oxidising domain and therefore the last step involves addition of subunits NDUFA12, NDUFS1, NDUFS4, NDUFS6, NDUFV1, NDUFV2 and NDUFV3, facilitated by NDUFAF2. In subsequent years this pathway has been revised with the

increased identification of assembly factors. Using BN-PAGE Andrews *et al.* explored the masses of the different subcomplexes and revised molecular mass estimations, in addition identifying TMEM126B and TIMMDC1 as assembly factors of the membrane arm, through siRNA knockdown that caused accumulation of low molecular mass subassemblies (Andrews *et al*., 2013).

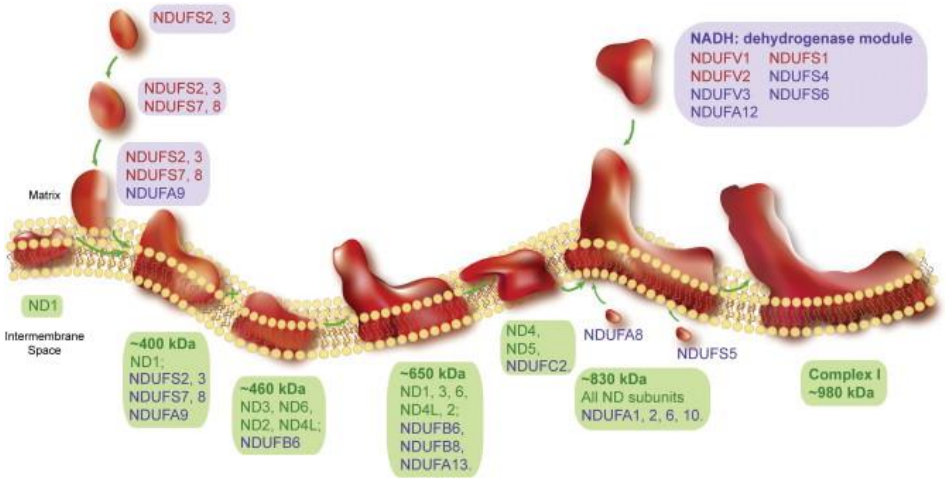


Figure 3.1 First assembly pathway of human complex I. This pathway was built around knowledge gained without a detailed structure of the enzyme and shown without assembly factors. The molecular masses of the subassemblies have since been revised by Andrews *et al.* and more assembly factors identified. Core nuclear subunits are labelled in red, mitochondrial core subunits in green and supernumerary subunits in blue. Figure adapted from Mimaki *et al.*, 2012.

There have now been 18 assembly factors of complex I identified– ACAD9, ATP5SL, COA1, ECSIT, FOXRED1, NDUFAF1-8, NUBPL, TIMMDC1, TMEM70, TMEM126B and TMEM186. Understanding of their function varies as shown in Table 3.1, although patient mutations demonstrate disruption of complex I assembly. Some have specific enzymatic activities, such as the hydroxylase NDUFAF5 and methylase NDUFAF7, whereas others have been shown to associate with membrane or matrix intermediates (Rhein *et al*., 2016 & 2015). In particular NDUFAF1, ECSIT and ACAD9 co-purify and co-migrate on BN-PAGE and have a role in ND subunit incorporation into the membrane arm subassemblies during mid-stage assembly (Nouws *et al.*, 2010).

Assembly factor	Homologues in other species	Stage of assembly	Description
NUBPL (IND1)		Early stage - *matrix arm associated*	Knockdown decreased the incorporation of iron-sulphur clusters and levels of the matrix arm were reduced (Sheftel *et al*., 2009).
ATP5SL		Early stage – *membrane arm associated*	Associated with complex I in NDUFA11 knockdown cells and in recent complexomics analysis of assembly (Andrews *et al*., 2013; Guerrero-Castillo *et al.*, 2017). Function unknown and controversial as an assembly factor.
TMEM70			Associated with membrane domain subcomplexes in complexomics analysis (Guerrero-Castillo *et al.*, 2017).
FOXRED1			FAD cofactor. Patient mutations result in complex I deficiency (Formosa *et al.*, 2015).
TIMMDC1 (C3orf1)		Early stage – *matrix arm associated*	Knockdown lowered cellular OCR and levels of intact complex I, with accumulation of subcomplexes (Andrews *et al*., 2013).
NDUFAF3 (C3orf60)			Patient mutations cause complex I deficiency and protein interacts with NDUFAF4 (Saada *et al*., 2009).
NDUFAF4 (C6orf66)			Patient mutations cause complex I deficiency (Saada *et al*., 2008).
NDUFAF5 (C20orf7)		Early stage – *matrix arm associated*	Arginine hydroxylase acting on NDUFS7 (Rhein *et al*., 2016).
NDUFAF6 (C8orf38)		Early stage - *matrix arm associated*	Stable knockdown decreased complex I levels and detectable ND1 in pulse translation experiments (Zurita Rendón & Shoubridge, 2012).
NDUFAF7 (C2orf56)		Early stage - *matrix arm associated*	Arginine methylase acting on NDUFS2 (Rhein *et al*., 2015).
NDUFAF8 (C17orf89)		Early stage – *matrix arm associated*	Interacts with and stabilises NDUFAF5. Knockdown in cell lines decreased complex I activity and lowered levels of NDUFAF5. Proposed candidate for complex I deficiency in a patient (Floyd *et al*., 2016).
NDUFAF1	CIA30 (*N. crassa*)	Mid-stage – *membrane arm associated*	Part of mitochondrial complex I assembly (MCIA) complex (Heide *et al*., 2012). Knockdown by siRNA decreases complex I by ~60% (Vogel *et al*., 2005). Patient mutations cause complex I deficiency (Dunning *et al*., 2007).
ACAD9			Component of MCIA complex. FAD cofactor. Knockdown and patient mutations cause complex I deficiency but no clear disruption of β-oxidation of fatty acids (Nouws *et al*., 2010).
ECSIT			Component of MCIA complex. Knockdown severely impaired complex I assembly with membrane arm intermediate accumulation and reduction in NDUFAF1 (Vogel *et al*., 2007).
TMEM126B			Component of MCIA complex. Knockdown lowered cellular OCR and levels of intact complex I with accumulation of subcomplexes (Andrews *et al*., 2013).
COA1 (C7orf44)		Mid-stage – *membrane arm associated*	Associated with membrane domain subcomplexes in complexomics analysis – possibly interacts with MCIA components (Guerrero-Castillo *et al.*, 2017). Complex IV assembly factor (Mick *et al*., 2012).
TMEM186 (C16orf51)		Mid-stage – *membrane arm associated*	Associated with membrane domain subcomplexes in complexomics analysis - possibly interacts with MCIA components (Guerrero-Castillo *et al.*, 2017).
NDUFAF2	13.4L (*N. crassa*)	Late stage – *matrix arm associated*	Result of gene duplication of NDUFA12. Non-essential in assembly (Pereira *et al*., 2013).

Table 3.1 Assembly factors of complex I. The eighteen proteins associated with complex I biogenesis are described. Some have defined functions whereas others are relatively newly identified and only demonstrated to interact with complex I subassemblies.

The most complete work on the assembly pathway to date harnessed a different approach for studying complex I subassemblies; going further than analysis of mass and protein identity from electrophoresis

gels (Guerrero-Castillo *et al.*, 2017). Nijtmans and co-workers used the complexomics approach, in which BN-PAGE is used to separate out complexes in their native form. Gel slices were then analysed by mass spectrometry to determine the distribution of protein complexes (Heide *et al*., 2012). Additionally, to follow complex I biogenesis the mitochondrial translation inhibitor chloramphenicol was used, to deplete mitochondria of assembled complex I: after removal of the inhibitor translation restarted, with samples taken at time points up to 24 hours for complexome profiling. The resulting scheme of assembly demonstrates many of the features previously identified – such as separate building of the membrane and matrix arms, and the involvement of assembly factors at defined stages – but also details further subcomplexes. Figure 3.2 shows that complex I assembles from four intermediates, the soluble N-domain, the connecting Q-domain, which contains the ubiquinone binding channel, and the membrane arm as P_P and P_D (proximal and distal membrane arm). The membrane arms elements were observed to be made up of four discreet collections of subunits: P_P-a, P_P-b, P_D-a and P_D-b and all the assembly factors are proposed to be released from the complex on addition of the N-domain. This data was used for comparison to complexomics undertaken in section 3.8 to enable assembly factor migration and the presence of subcomplexes to be assessed.

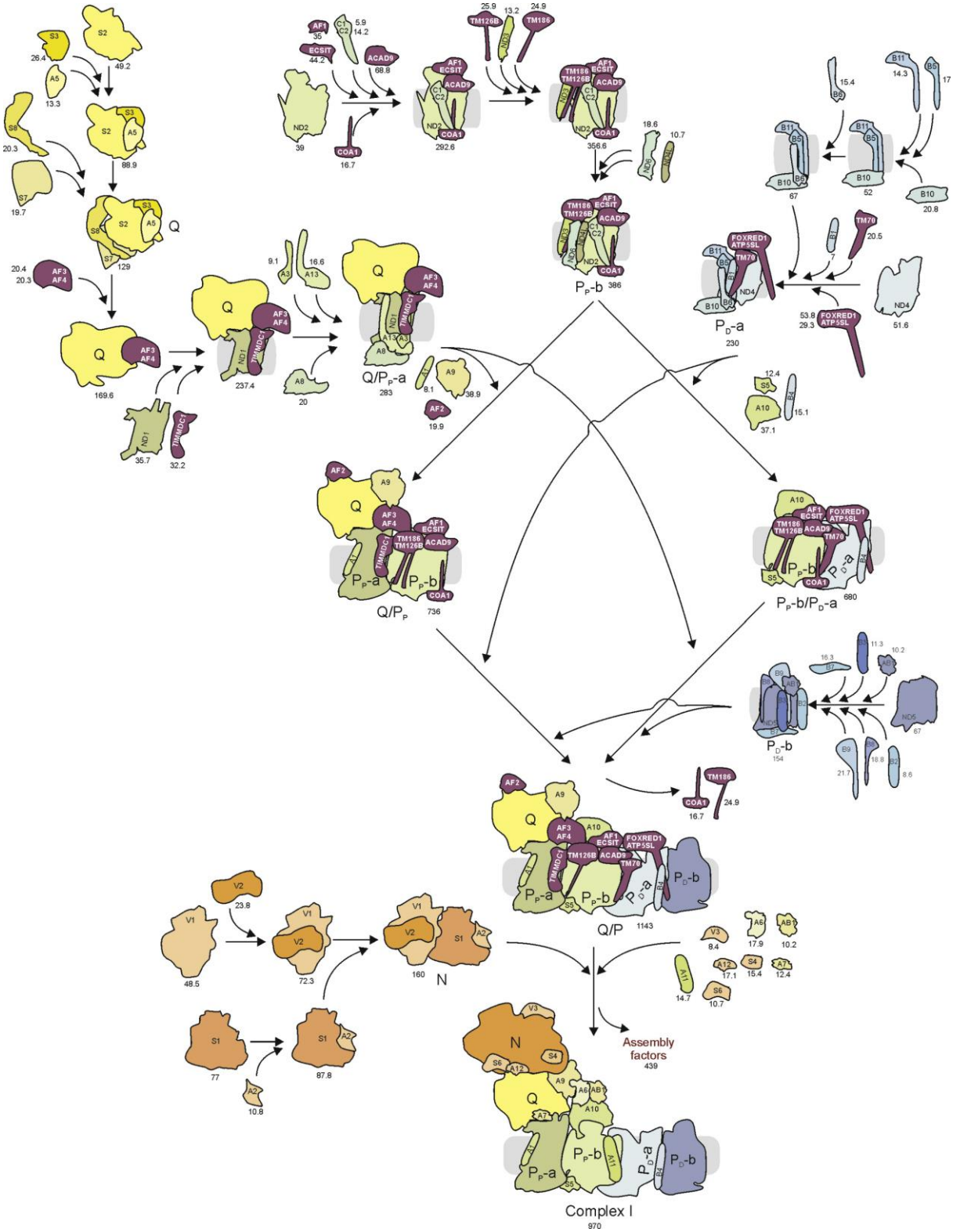


Figure 3.2 Assembly pathway of complex I determined using the complexomics approach. The most recent pathway for complex I assembly utilised chloramphenicol to inhibit mitochondrial translation and remove complex I from mitochondria. Upon removal of the inhibitor complex I biosynthesis resumed and time points up to 24 hours were assessed by complexomics to track the assembly of the enzyme. The identities of different intermediates are shown with their calculated masses. Subunits are coloured by their structural module: N-domain in orange, Q-domain in yellow, P-$_P$ in green and P-$_D$ in blue. Assembly factors are shown in purple. All 17 assembly factors detailed in Table 3.1 were observed and 13 were found associated with complex I subcomplexes. Figure from Guerrero-Castillo *et al.*, 2017.

3.1.2. Flavin insertion into the NDUFV1 subunit of complex I

Figure 3.2 shows that understanding of complex I assembly is becoming increasingly detailed, especially with the assignment of assembly factors to particular steps. To catalyse NADH oxidation and quinone reduction the enzyme requires a series of cofactors: FMN and a chain of iron sulphur clusters for electron transport. NUBPL has been assigned as an iron sulphur cluster assembly factor: knockdown caused both accumulation of membrane domain subcomplexes and a decrease in the amount of iron associated with complex I (Sheftel *et al*., 2009). Despite this, no assembly factor has been found for incorporation of FMN into NDUFV1. An assembly factor may be required as the flavination of the complex II subunit SDHA requires a designated chaperone, SDHAF2, first identified in yeast (Hao *et al*., 2009). However, FAD is covalently bound to SDHA whereas in NDUFV1 FMN is non-covalently bound, which may dictate the insertion method (Starbird *et al*., 2015). NDUFV1 is assembled onto the growing enzyme in the very late stages, as part of the N-domain (orange in Figure 3.2). However, it is not known whether the full enzyme would still assemble without FMN bound to NDUFV1 and how overall stability would be affected, therefore this remains an aspect of assembly that has yet to be explored.

3.1.3. Cellular flavin cofactor synthesis

The cofactors FAD and FMN are derived from riboflavin, vitamin B2, which must be obtained by dietary uptake. Riboflavin is absorbed from the small intestine, passed into the blood stream and accumulated by cells; each stage requires a specialised transporter to facilitate movement across cell membranes (Yonezawa *et al*., 2008). This pathway, although crucial for providing the cofactor required by many proteins, is only partially defined. Cytosolic riboflavin kinase (RFK) has been characterised; it phosphorylates riboflavin using ATP to form FMN (Karthikeyan *et al.*, 2003). Mitochondrial localisation of RFK has also been suggested, however there is currently no experimental evidence to substantiate this (Barile *et al*., 2000). FAD synthase (FADS) then converts FMN to FAD. Importantly, FADS was known to be present in the cytoplasm but more recently a human isoform has been located in mitochondria (FADS1) by Barile and co-workers; it differs from the cytoplasmic isoform (FADS2) by an additional, predicted, 97 amino acid mitochondrial target sequence. Additionally, FADS1 was shown to be imported by rat liver mitochondria and fluorescent labelling in mammalian cells also supported its mitochondrial localisation (Torchetti *et al.*, 2010).

The plasma membrane riboflavin transporter, SLC52, was identified after the clinical observation that some complex I disorders were effectively treated by dietary riboflavin supplementation and mutations were mapped to a transporter family (Chiong *et al*., 2007*;* Yonezawa *et al.*, 2008). The three members of the SLC52 gene family: SLC52A1, A2 and A3 display differing tissue expression; it is suggested that

while A2 is highly expressed in the brain, A1 and A3 coordinate riboflavin uptake from the small intestine lumen into the blood, and also cellular uptake of the vitamin (Yonezawa *et al.*, 2013).

Many flavoproteins reside within organelles, including complex I in mitochondria, and so there must be transport of flavins across the IMM. FLX1 has been assigned the flavin carrier protein in yeast; it was discovered through complementation of respiratory deficient *Saccharomyces cerevisiae* strains **(**Tzagoloff *et al.*, 1996). A human orthologue for FLX1 was found to rescue growth of a *S. cerevisiae* FLX1 mutant, which suggested that this protein, SLC25A32, was the mitochondrial FAD transporter (Spaan *et al.*, 2005). However, this protein was previously characterised as a folate transporter (SLC25A32) since growth rescue of a glycine auxotrophic Chinese hamster ovary cell line when transfected with SLC25A32 increased mitochondrial folate levels (Titus *et al*., 2000). It may be possible that due to structural similarities of flavin and folate head groups, both pteridine ring-derived, substrate promiscuity could allow the transport of multiple substrates by a single protein (Spaan *et al.*, 2005).

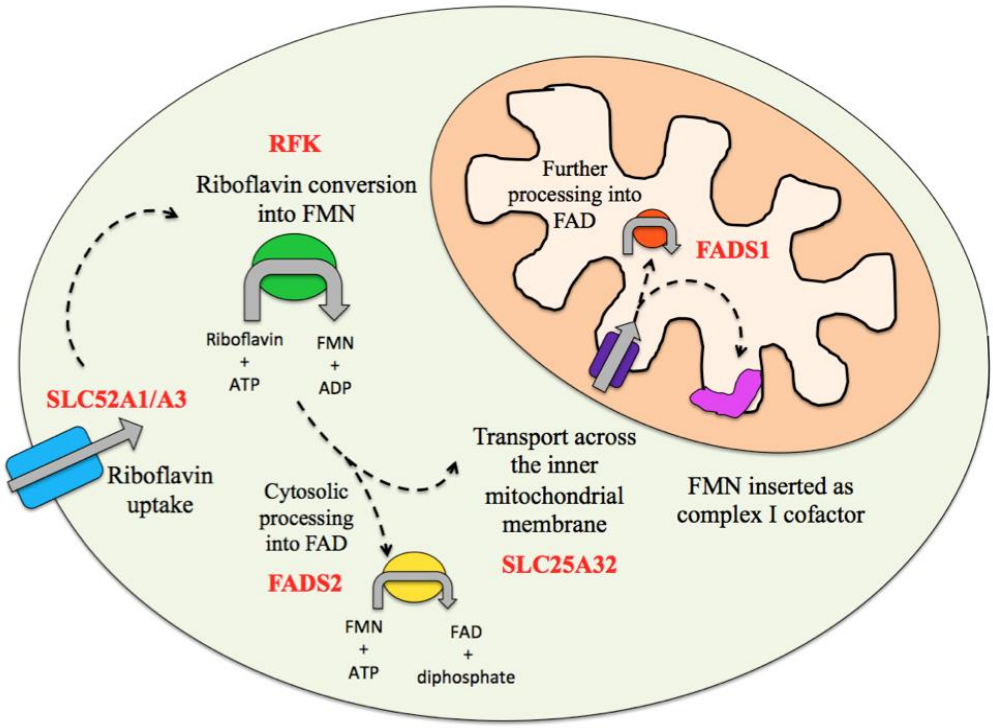


Figure 3.3 Schematic diagram of the proposed riboflavin pathway. Using evidence from previous research a putative riboflavin uptake and processing pathway was constructed: vitamin B2 is first transported into the cell (mint) across the plasma membrane by SLC52A1/A3 (blue), where it is converted into FMN by RFK (green) in the cytosol. FADS2 (yellow) converts FMN into FAD in the cytosol. The putative riboflavin transporter SLC25A32 (purple) carries FMN into the mitochondrial matrix (cream) across the IMM. FMN is then converted into FAD by FADS1 (orange) or inserted into NDUFV1 subunit of complex I (magenta).

Current knowledge of flavin processing within the cell is summarised in Figure 3.3. Here, to investigate FMN incorporation into complex I three gene candidates were chosen – RFK for its role in FMN production and SLC25A32 for its proposed mitochondrial flavin uptake. Additionally, since this project was started, riboflavin responsive mutations in SLC25A32 have been reported that further suggest its role in mitochondrial FAD transport (Hellebrekers *et al*., 2017). However due to lack of an enzyme producing FMN in the mitochondria it is likely that a transporter is at least required for FMN transport, as FADS1 is present to generate FAD in the matrix. FADS1 was also selected due to a possible indirect effect on complex I assembly - the assembly factors ACAD9 and FOXRED1 require FAD - therefore it may provide further insight into assembly of the enzyme.

3.2. Aims

The aims of work in this chapter were to investigate how FMN is incorporated into complex I, specifically to:

- Identify subcomplexes and assembly factors concerned with FMN insertion into NDUFV1.
- Understand the direct effects of depleting FMN on the activity and assembly of complex I.
- Further characterise the upstream pathway: particularly mitochondrial riboflavin uptake and modification into flavin cofactors.
- Investigate the cellular and mitochondrial effects of riboflavin depletion.

3.3. The importance of riboflavin in the growth medium and impact on cell line investigations

3.3.1. Initial protein knockdown experiments with siRNA

To restrict FMN incorporation into complex I three proteins of the riboflavin-processing pathway were targeted. RFK represents one of the best-characterised enzymes and the mitochondrial isoform of FADS was also selected, FADS1, because several complex I assembly factors require FAD cofactors. The mitochondrial transporter SLC25A32, which has a putative folate transporting function, was also chosen as it has been linked with flavin transport. This area of the pathway still lacks understanding; therefore studying the transporter may increase our functional understanding through comparison with the RFK and FADS1 knockdown effects.

For each gene three siRNA oligonucleotides were designed for knockdowns in 143B CyWT cells (Chapter 2, Table 2.1). Transfection was undertaken as described (Chapter 2, section 2.3.1) and after a 72-hour incubation with 50 nM siRNA, the level of protein expressed was determined using Western Blotting. Transfection resulted in some cell death, however the surviving 143Bs grew well and achieved sufficient cell density for OCR analysis. The effect on 143B mitochondrial respiration was then analysed with a stress test, using a XF96 extracellular flux analyser (Seahorse Biosciences), described in Chapter 2, section 2.6.1. Western Blot analysis demonstrated that, for all three genes, siRNA treatments lowered relative expression of the target protein, notably siRNA-1 for SLC25A32, siRNA-1 and -2 for RFK and siRNA-1 for FADS1 at 50 nM (Figures 3.1, 3.2 and 3.3).

Despite efficient siRNA knockdown of all three proteins (Figures 3.4A, 3.5A and 3.6A) none of the 143Bs displayed a respiratory phenotype when compared to transfection agent-treated 143B controls (B in the respective figures).

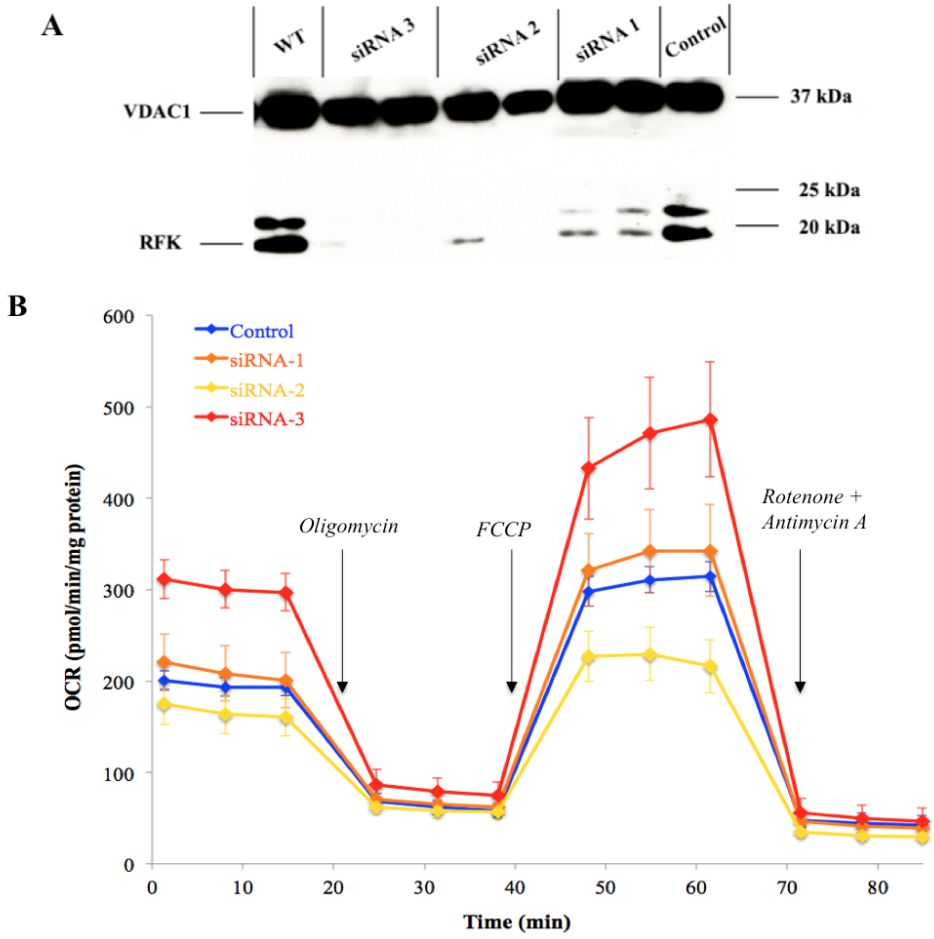


Figure 3.4.A Efficiency of RFK siRNA knockdown assessed by Western Blot. 20 µg cell lysate samples from siRNA transfected cells and non-transfected 143Bs were blotted for RFK (17.6 kDa, antibody from Proteintech) to determine the level of knockdown (duplicate wells). Two controls were used, non-treated 143B lysate (WT) along with Lipofectamine™and Opti-MEM treated 143B lysate (control), to account for effects caused by transfection reagent treatment. The membrane was also blotted for VDAC1 (37 kDa, antibody from Abnova), demonstrating slight differences in loading. **B. Seahorse extracellular flux data comparing the OCR of control 143Bs to each knockdown.** Cells were seeded at 30,000 cells/well on the morning of the assay to minimise the time that transfected cells were incubated in siRNA-free medium. 143B cells treated with Opti-MEM and Lipofectamine™ 2000 were used as the control (blue diamonds). 16 wells of each condition were assayed and the experiment repeated thrice (mean results for one representative assay shown, with error bars denoting standard error) with siRNA-1 in orange, siRNA-2 yellow, siRNA-3 red.

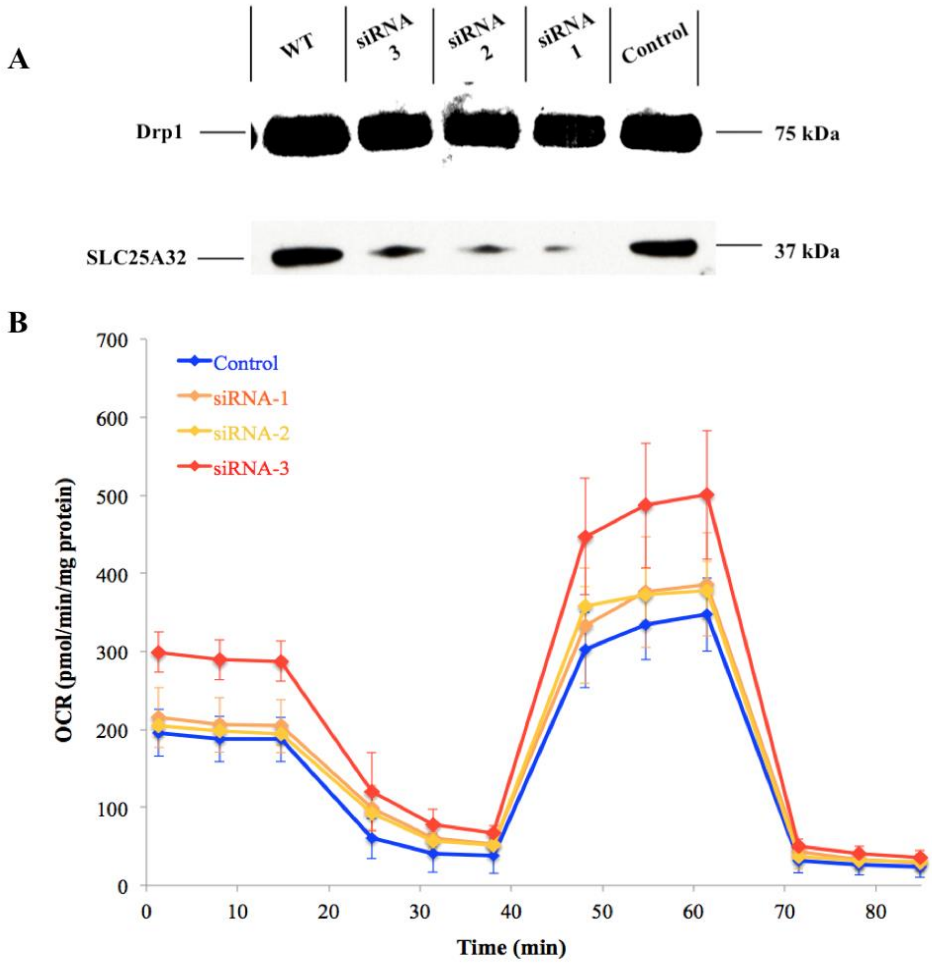


Figure 3.5A. Efficiency of SLC25A32 siRNA knockdown assessed by Western Blot. 20 μg cell lysate samples from transfected cells and non-transfected 143Bs were blotted for SLC25A32 (35.4 kDa, antibody from Abcam) to determine the level of knockdown with each siRNA. Two controls were used, non-treated 143B lysate (WT) along with Lipofectamine™and Opti-MEM treated 143B lysate (control), which was treated with the same volumes of Opti-MEM and Lipofectamine™ 2000 used to transfected the siRNAs. To ensure loading was consistent the membrane was also blotted for Drp1 (75 kDa, Abnova), demonstrating even loading. **B. Seahorse extracellular flux data comparing the OCR of control 143Bs to each knockdown.** Cells were seeded at 30,000 cells/well on the morning of the assay to reduce the time that transfected cells were incubated in siRNA-free medium. 143B cells treated with Opti-MEM and Lipofectamine™ 2000 were used as the control (blue). 16 wells of each condition were assayed and the experiment repeated thrice (mean results for one representative assay shown, with error bars denoting standard error) with siRNA-1 in orange, siRNA-2 yellow, siRNA-3 red.

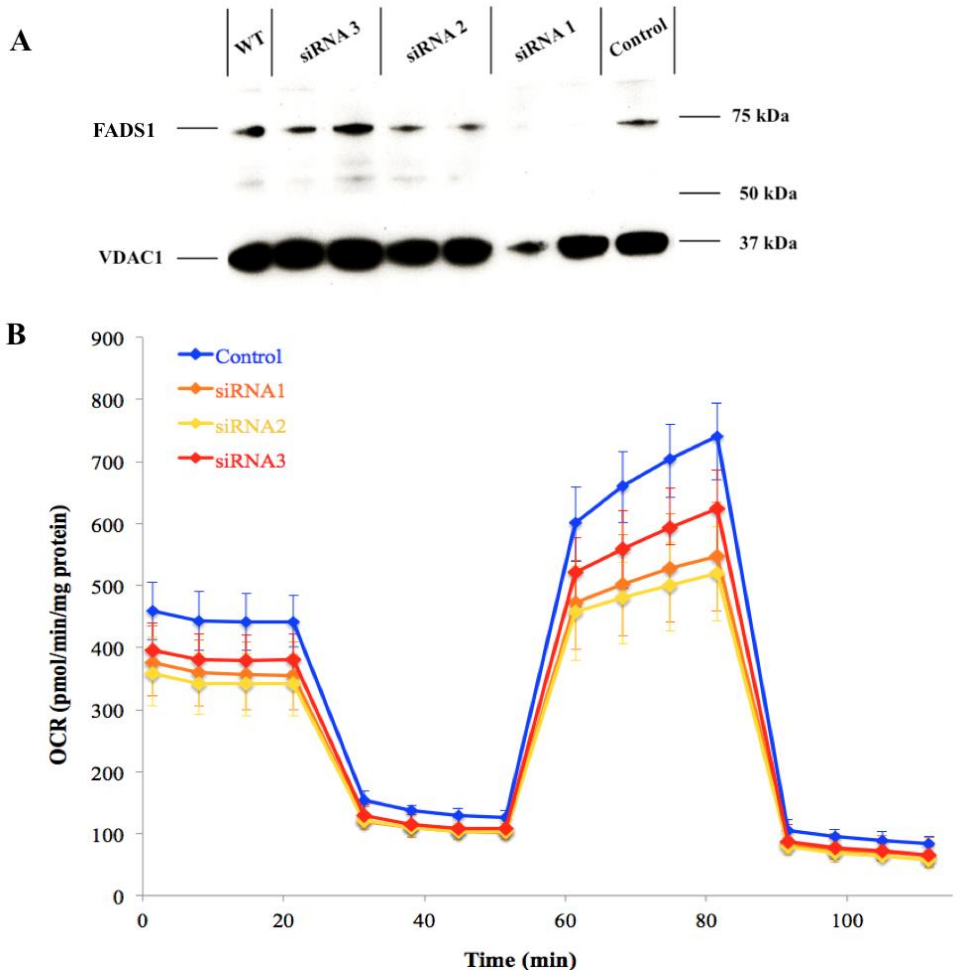


Figure 3.6A. Efficiency of FADS1 siRNA knockdown assessed by Western Blot. 10 µg cell lysate samples from transfected cells and non-transfected 143Bs were blotted for FADS1 (65.3 kDa, antibody from Aviva Systems Biology) to determine the level of knockdown with each siRNA (duplicate wells). Two controls were used, non-treated 143B lysate (WT) along with Lipofectamine™and Opti-MEM treated 143B lysate (control), to account for effects caused by transfection reagent treatment VDAC1 (37 kDa) was used as a loading control, demonstrating even loading except for one duplicate. **B. Seahorse extracellular flux data comparing the OCR of control 143Bs to each knockdown.** Cells were seeded at 30,000/well on the morning of the assay to reduce the time that transfected cells were incubated in siRNA-free medium. Control cells (blue) were assayed, with 16 wells of each condition and the experiment was repeated thrice (mean results for one representative assay shown, with error bars denoting standard error). siRNA-1 in orange, siRNA-2 yellow, siRNA-3 red.

Two bands observed around the expected molecular weight for RFK displayed similar knockdown levels; these may reflect different isoforms indicated in the Ensembl database (Zerbino et al., 2018). Although knockdown was effective there was no meaningful change in OCR for SLC25A32 or RFK compared to a non-transfected lipofectamine control (Figure 3.4, 3.5B). The results suggested that cells maintained FMN insertion into complex I and there was catalytic activity even when the abundance of RFK was diminished. The same observation was made for SLC25A32. It should be noted that some siRNA treatments (siRNA-3, red, in Figures 3.4B and 3.5B) resulted in OCRs above the control – this variability, despite similar knockdown levels with the three siRNA for RFK and SLC25A32, also undermined the reliability of this data. Nevertheless, the control data were not consistently higher than the knockdowns. For FADS1 the level of knockdown varied; siRNA-3 showed little change in protein expression whereas siRNA-1 resulted in almost complete knockdown, Figure, 3.6A. Unfortunately all available antibodies targeted all isoforms and so, whilst the siRNA sequences chosen were specific to FADS1, blotting showed the levels of both FADS isoforms. Measurement of OCR showed decreased basal respiration of all knockdown cells and some correlation with the observed level of knockdown; siRNA-3 showed least effect whereas siRNA-1 and -2 (Figure 3.6B) resulted in basal rate decrease of ~25% compared to the control. However, in replicate experiments, the difference between knockdown and control base rates was variable, with increased error and weaker correlation with protein abundance.

To further investigate the possibility of an effect, a dual knockdown was tested; either siRNA-1 or the siRNA-2 oligonucleotides for both RFK and FADS1 were co-transfected at 100 nM, to increase the knockdown efficiency. The levels of knockdowns achieved are shown in Figure 3.7A and indicate that co-transfection decreased expression of both RFK and FADS1. However, there was still no decrease in OCR in the double knockdown; as noted previously some experiments even showed a higher basal OCR for knockdowns (siRNA-1 combination, green, in Figure 3.7B).

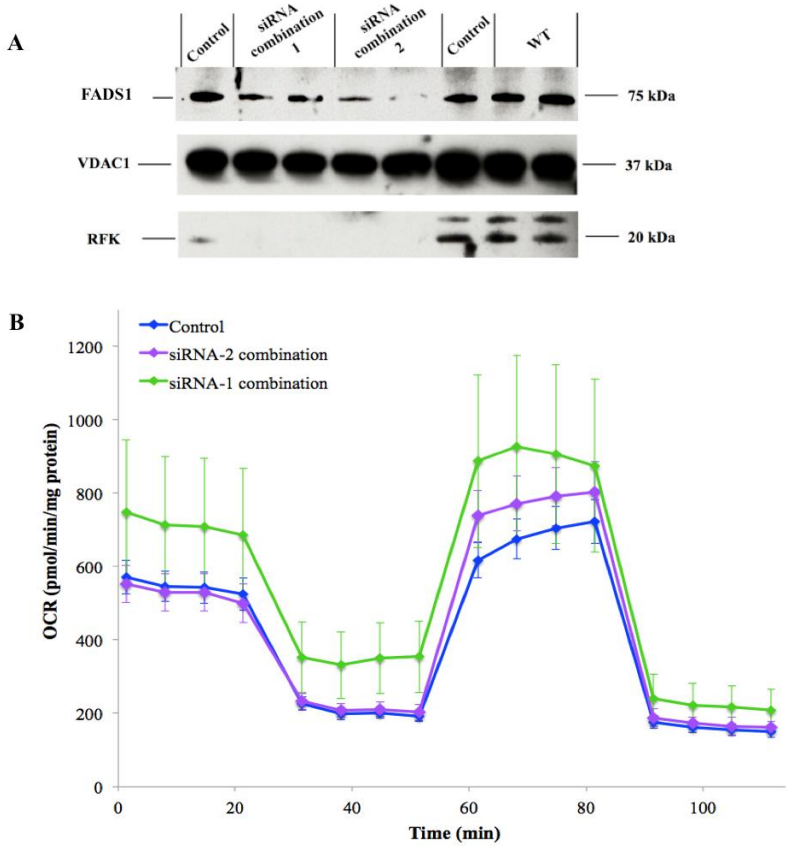


Figure 3.7A Western Blot to assess the efficiency of the dual RFK and FADS1 siRNA knockdowns at 100 nM. 10 µg cell lysate samples from transfected cells and non-transfected 143Bs were blotted for FADS1 (65.3 kDa) and RFK (16.4 kDa) to determine the level of knockdown with combinations of siRNA-1 for RFK and FADS1 or siRNA-2 for RFK and FADS1 (duplicate wells). To ensure loading was consistent the membrane was also blotted for VDAC1 (37 kDa), demonstrating slightly higher loading of control and WT samples. **B. Seahorse extracellular flux data comparing the OCR of control 143Bs to each knockdown.** Cells were seeded at 30,000/ml on the morning of the assay to reduce the time that transfected cells were incubated in siRNA-free medium. 24 wells of each condition were assayed and the experiment was repeated thrice (mean results for one representative assay shown, with error bars denoting standard error) with control cells (blue), siRNA combination 1 (green) and siRNA combination 2 (purple).

Lack of respiratory phenotype for all three protein knockdowns suggested that, while protein expression was lowered, the amount of FMN produced was still sufficient for normal incorporation into complex I. Therefore, from these preliminary assays alone it was not possible to confirm the roles of these proteins in riboflavin handling or gain further understanding of FMN incorporation into complex I.

3.3.2. Targeting riboflavin uptake to decrease cytosolic riboflavin

The lack of a distinct phenotype across all knockdowns suggested that the riboflavin content of the growth medium (DMEM) was compensating for any restriction in the processing ability of the cells. Standard DMEM contains 1 μM riboflavin (data from supplier) and 10% FBS contributed an additional 0.77 nM (Said *et al*., 1998). Therefore a number of approaches were tested to lower the availability of riboflavin and limit the cellular supply for flavin cofactor synthesis.

3.3.2.1. Uptake inhibition with lumiflavine

Members of the plasma membrane riboflavin transporter family, SLC52, have been characterised and their expression investigated (Yonezawa et al., 2008; Yoshiaki et al., 2010). SLC52A1 and A3 have been implicated in riboflavin absorption and cellular uptake, however which transporter is expressed in different tissues remains ambiguous (Yonezawa et al., 2013). Thus, rather than targeting individual transporter types with siRNA, an inhibitor was chosen to decrease overall riboflavin uptake. Lumiflavine is a riboflavin analogue that competes for the substrate-binding site and has previously been used for direct uptake assays. In published experiments, lumiflavine was used for short-term incubation and no investigation was made of the downstream consequence of diminished cytosolic riboflavin. Ma and co-workers measured radiolabelled riboflavin uptake in ARPE-19 cells, with addition of structural analogues and related compounds to determine the specificity of uptake (Said et al., 2005). 1 μM lumiflavine decreased uptake of 40 nM ^{3}H-riboflavin by ~50%, increasing in a concentration dependent manner - at 10 μM uptake was lowered to ~3%. Additionally, Said et al measured the concentration dependent inhibition of 5 nM ^{3}H-riboflavin uptake by HepG2 cells and calculated an inhibition constant (K_i) of 1.84 μM after a 3-minute incubation (Said et al., 1998). Although this work and other uptake studies involved low extracellular riboflavin concentrations it suggested that lumiflavine could be used to inhibit riboflavin uptake. Consequently 1 μM, 5 μM and 10 μM lumiflavine additions to standard DMEM (containing 1 μM riboflavin) were used and cells incubated for 72 hours before the OCRs were measured.

There was no significant difference in respiratory profile for the cells incubated in lumiflavine observed when compared to control cells, Figure 3.8. Treatment with 20 μM lumiflavine also showed no change.

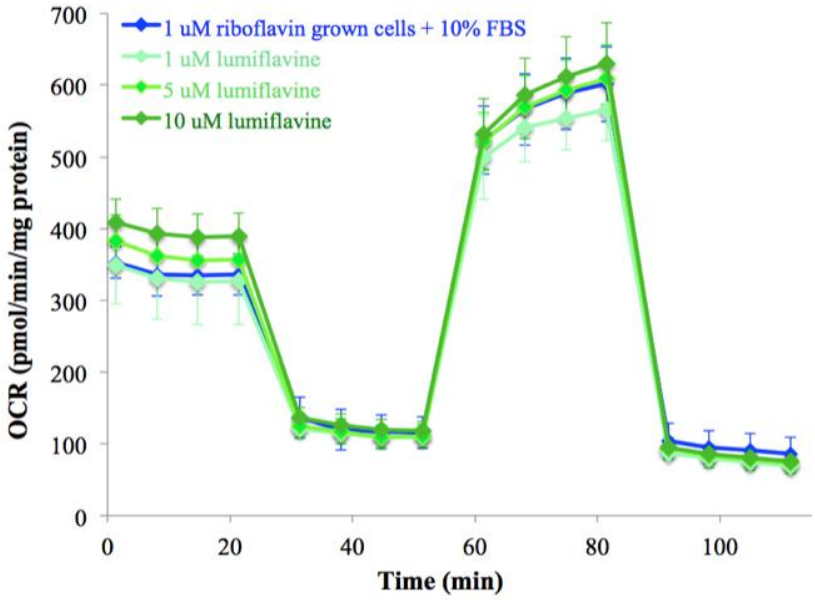


Figure 3.8 Seahorse extracellular flux data comparing the OCR of control 143Bs to the OCR of 143Bs treated with lumiflavine. Cells were seeded at 11,500/well the day before the assay and 24 wells of each condition were assayed, with the experiment repeated twice (mean results for one representative assay shown, with error bars denoting standard error). Control cell data is shown in blue, 1 μM lumiflavine-treated in light green, 5 μM lumiflavine-treated in bright green and 10 μM lumiflavine-treated in dark green.

3.3.2.2. Decreasing FBS content in the growth medium to lower riboflavin availability

Riboflavin deficiency has been studied in cultured cells, predominantly by Zempleni and co-workers, with concentrations based upon earlier work on human plasma riboflavin concentrations. 532 nM riboflavin was a sufficient concentration whereas 3.1 nM reflected riboflavin deficiency (Werner et al., 2005). Notably after pre-treatment of HepG2 cells (5 weeks on 12.6 nM riboflavin) lowering riboflavin concentration to 3.1 nM induced cell cycle arrest, >90% decrease in flavin-dependent enzyme activity (measured by glutathione reductase) and <20% expression of RFK and FADS1 mRNA after 4 days (Werner et al., 2005). In light of this work, 143B cells were grown in riboflavin-free DMEM, supplemented with different volumes of standard DMEM, to control the riboflavin concentration. After at least 72 hours in 500 nM, 200 nM or 10 nM riboflavin-containing medium the OCR was measured and compared to that from 143Bs grown in 1 μM riboflavin. There was no observable morphology change or difference in growth rate and no significant difference in respiration rates (data not shown). Therefore, the concentration of riboflavin was further depleted to 4.77 nM (0.77 nM contributed by 10% FBS supplementation) for another 72 hours; again there was no observable change in OCR (Figure 3.9).

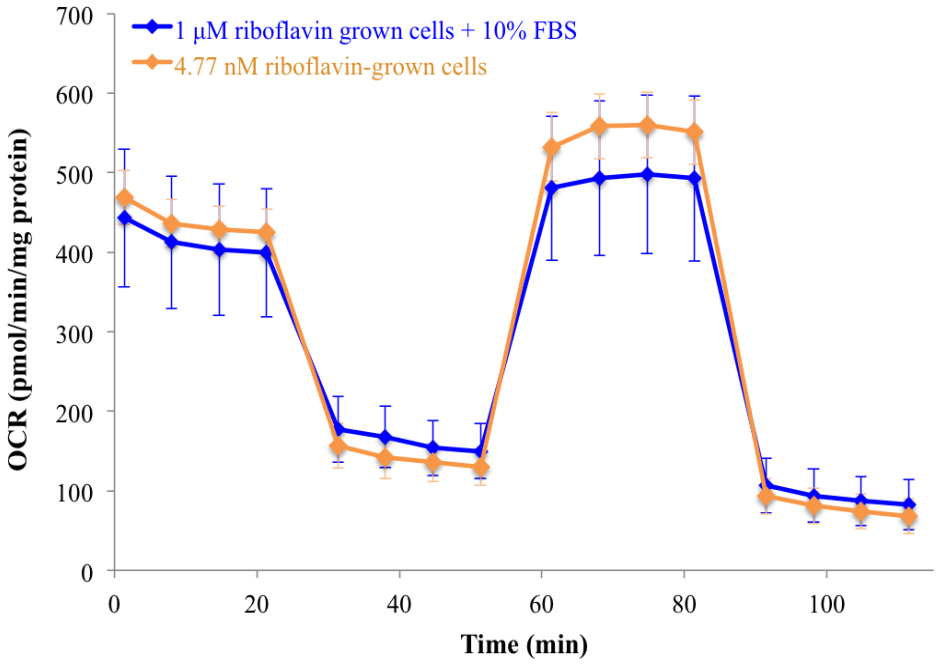


Figure 3.9 Seahorse extracellular flux data comparing the OCR of control 143Bs cells grown in riboflavin depleted medium. Cells were seeded at 11,500/well the day before the assay and 48 wells of each condition were assayed with the experiment repeated twice (mean results of one representative assay shown with error bars denoting standard error). Control cell data are shown in blue and data for cells incubated in 4.77 nM riboflavin in orange.

Further depletion of extracellular riboflavin (4.77 nM, 3.77 nM, 2.77 nM, 1.77 nM and 0.77 nM riboflavin containing DMEM) for 1 week also matched the phenotype of 1 µM riboflavin-grown cells. Increasing the incubation time to 9 days also showed comparable OCRs to the 1 µM riboflavin-grown cells (data not shown). These results indicate the importance of the extended low riboflavin pre-treatment stage adopted by Zempleni and co-workers.

3.3.3. Extended depletion of riboflavin is necessary to promote a phenotype

Cells were continually cultured in 0.77 nM riboflavin (riboflavin free DMEM + 10% FBS) for one month, to limit the cellular supply of riboflavin but maintained the cells in a relatively stress free state as there were no observable phenotypic changes. These cells are referred to as ‘conditioned’ cells and were used as the basis for preparing samples for improved investigation into changes in FMN incorporation into complex I.

First, the use of synthetic, riboflavin free serum was tested by progressively introducing it to replace FBS. However, complete replacement with synthetic serum was not tolerated by 143B cells so could not be used to decrease riboflavin availability.

Second, conditioned 143Bs were treated with 50 μM lumiflavine for 3 days, but despite initial growth they failed to adhere when seeded in 96 well plates for OCR analysis; therefore reliable respiratory measurements could not be obtained. The concentration of lumiflavine was thus decreased to 35 μM but again the cells responded poorly. 20 μM lumiflavine treatment with a 7 day incubation resulted in good growth of cells with a distinct morphology, and a significant change in OCR. However, these cells were not viable beyond one week.

3.3.3.1. Riboflavin-free DMEM with decreased FBS additions

Supplementation of riboflavin-free DMEM with 10% FBS contributed only 0.77 nM riboflavin yet no phenotype was observed. Therefore lowering FBS addition further was used to decrease riboflavin levels and a series of different FBS levels were tested on conditioned cells (for at least one week). After 8 days those grown in 2% FBS-containing riboflavin-deficient DMEM (0.15 nM riboflavin from FBS) were assayed. A significant decrease in basal respiration rate was observed for 0.15 nM riboflavin-grown cells, by over 50%, although no change was observed regarding FCCP-induced respiration or non-mitochondrial, rotenone insensitive respiration (data not shown).

To ensure this effect was due to the low concentration of riboflavin, rather than a decrease in growth-promoting factors present in FBS, a further control was included, 1 μM riboflavin-containing medium +2% FBS (light blue Figure 3.10 and 3.11). After 5 days in 0.15 nM riboflavin or 1 μM riboflavin +2% FBS the OCRs were measured alongside 143B cells grown in 1 μM riboflavin +10% FBS. Figure 3.11 confirms that the lower basal OCR measured in 0.15 nM riboflavin-grown cells was due to the further lowering of extracellular riboflavin concentration; both controls showed a comparable respiration profile whereas the 0.15 nM cells displayed a ~20% decrease in basal OCR in this example. Figure 3.10 demonstrates that a shift in metabolism occurred; 0.15 nM riboflavin-grown cells displayed increased basal ECAR but lowered OCR, suggestive of a glycolytic phenotype. The difference in basal OCR was less distinct than after 8 days of treatment, suggesting a progressive nature to the phenotype, which was then investigated further.

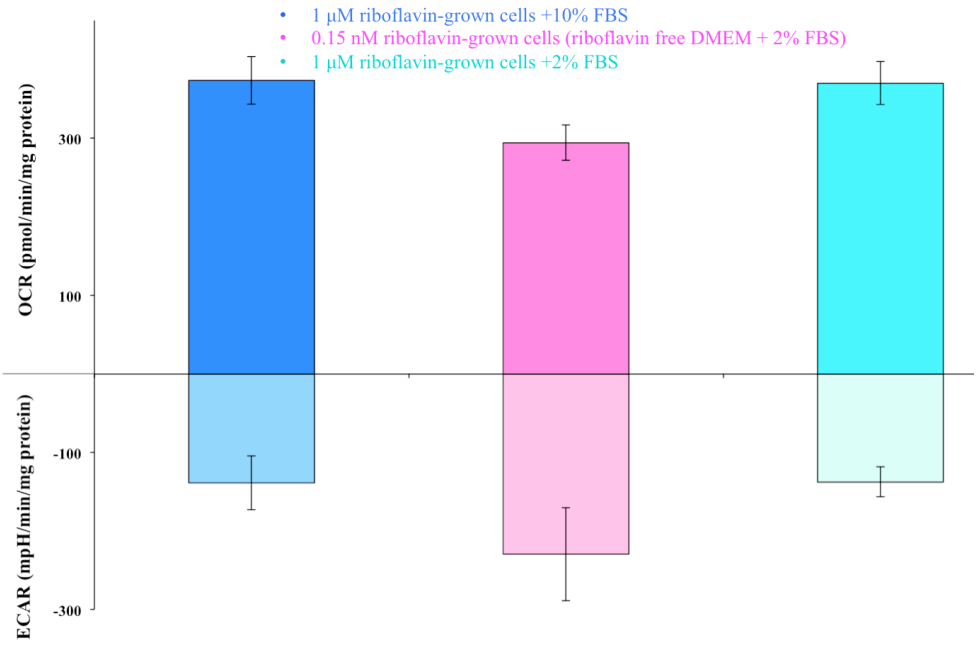


Figure 3.10 Seahorse extracellular flux data comparing the basal OCR and ECAR of control 143Bs to OCR of cells grown in riboflavin-depleted medium. After passaging for 4 days cells were seeded at 11,500/well the day before the assay. 40 wells of 0.15 nM riboflavin-grown cells (pink) and 1 µM riboflavin +2% FBS cells (light blue) were assayed along with 12 wells of 1 µM riboflavin +10% FBS treated cells (dark blue). Mean basal OCR is presented as positive values whereas mean ECAR on the same axis has negative values (with standard error) to show the relative split between oxidative phosphorylation and glycolytic pathways between the different cells. Data from one representative assay are shown.

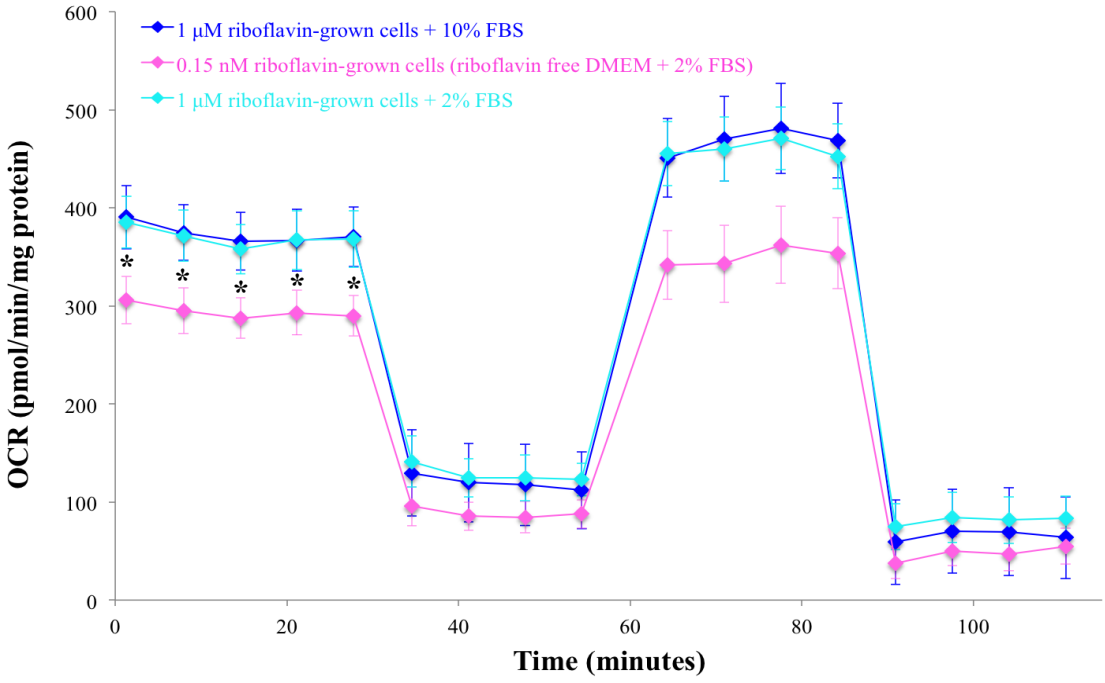


Figure 3.11 Seahorse extracellular flux data comparing the OCR of control 143Bs to OCR of cells grown in riboflavin-depleted medium. After passaging for 4 days cells were seeded at 11,500/well the day before the assay, 40 wells of 0.15 nM riboflavin-grown cells (pink) and 1 µM riboflavin +2% FBS (light blue) were assayed along with 12 wells of 1 µM riboflavin +10% FBS treated cells (dark blue). Error bars denote standard error and asterisks show that the basal OCR of riboflavin deficient cells was significantly different from both control cell samples (which were not significantly different to one another), calculated using unpaired student t-tests, $P < 0.05$. Data from one representative assay are shown.

143B cells cultured in 0.15 nM riboflavin for >5 days developed a distinct morphology: they shifted from their usual epithelial-like morphology to more spindle-like, almost fibroblast morphology (Figure 3.12); consistent with the change in OCR shown in Figure 3.11. This change was monitored using Incucyte software for 160 hours to compare the growth rate and physical characteristics with 1 µM riboflavin-grown cells supplemented with either 10% or 2% FBS. Figure 3.13 shows the growth curves of each sample, with cells grown in 1 µM riboflavin-containing DMEM +10% FBS reaching ~100% confluence after 100 hours (dark blue). Cells cultured in lower FBS (+2%) but 1 µM riboflavin-containing DMEM displayed slower growth with a longer lag phase, however after ~148 hours confluence was achieved (light blue). In contrast, the 0.15 nM riboflavin-grown cells did not proliferate above ~70% confluence and growth plateaued after 100 hours (pink). The erratic nature of the final data points is a result of increased cell death (rounded-ended arrow in Figure 3.14K), which interfered with the density calculation, highlighted in the images of Figure 3.14K and L.

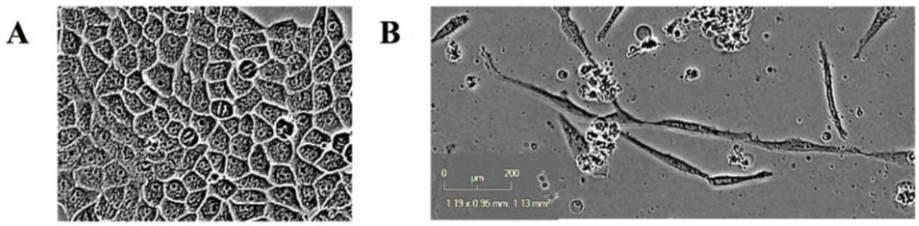


Figure 3.12 143Bs grown in 0.15 nM riboflavin display a distinct morphology from cells grown in 1 µM riboflavin. Cells were seeded at 25,000/well with 2.5 ml of appropriate medium and incubated at 37ºC for 160 hours. **A.** Incucyte software was used to record the change in growth: control 143B cells grown in standard DMEM showed normal epithelial-like morphology, ~100 µm in diameter. **B.** Conditioned 143Bs grown in 0.15 nM riboflavin demonstrated a change in cell shape, becoming elongated with a cell length ~300 µm.

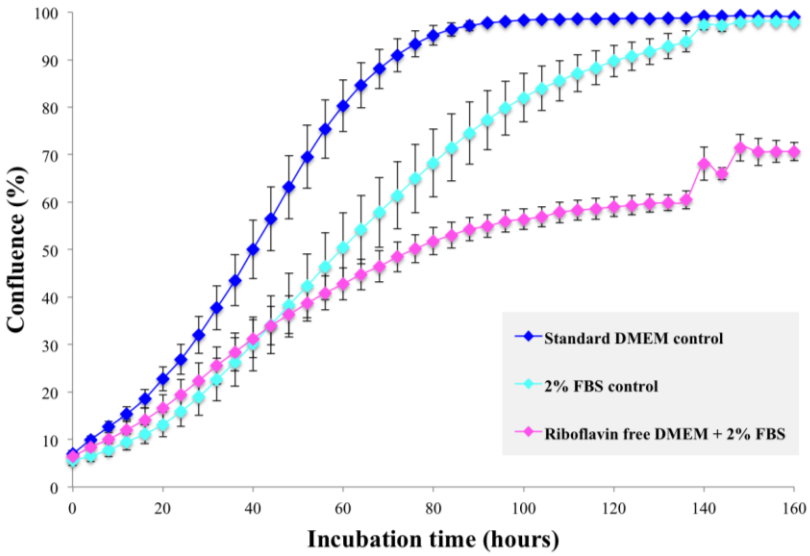


Figure 3.13 Growth profile of cells grown under different riboflavin concentrations. Cells were seeded at 25,000/well with 2.5 ml of appropriate medium and incubated at 37 ºC for 160 hours. Incucyte software was used to record the change in growth, with images obtained every 2 hours and percentage confluence calculated with standard error. Two controls were used: 1 µM riboflavin +10% FBS in dark blue and 1 µM riboflavin +2% FBS in light blue to compare with the growth and morphology of 0.15 nM riboflavin-grown cells (pink). Data from one representative assay are shown.

Figure 3.14 shows the morphology of the two controls (1 μM riboflavin +10% FBS or 2% FBS) were similar, although the +2% FBS cells showed slower growth and some cell elongation (arrows) when cell density was low (Figure 3.11E and F). Figure 3.14I-L demonstrates the distinct morphology of cell grown in 0.15 nM riboflavin; cells were elongated (arrows) and a high proportion underwent cell death (round-ended arrow) – with a mesh-like cell network compared to the tight packing of the control cells. This supports the OCR data that suggested the phenotype observed was progressive; cells either adapted to the greatly depleted riboflavin concentrations with altered basal respiration rate, or died. Additionally, this experiment highlighted that reducing cellular riboflavin by depleting riboflavin in the medium had a widespread effect on cell growth and morphology: alterations caused by general riboflavin deficiency were not limited to complex I alone.

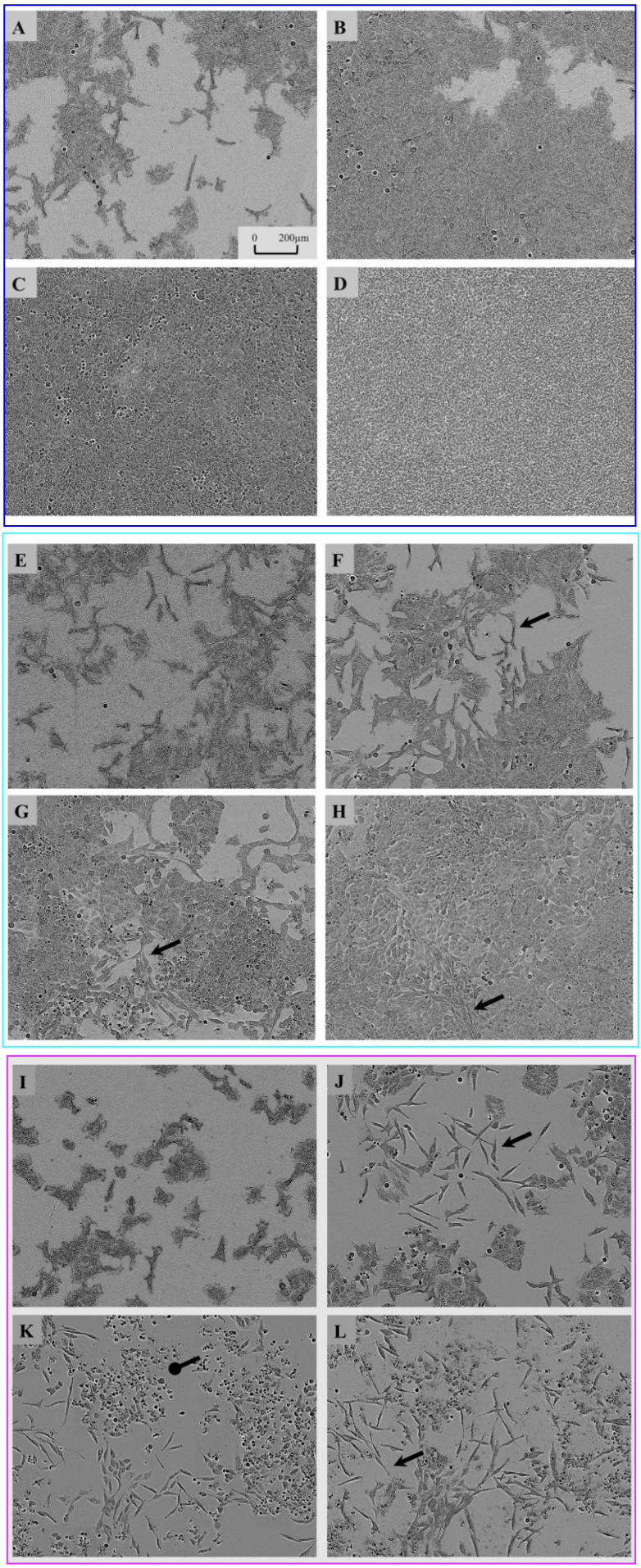


Figure 3.14 Time interval cell growth images to compare riboflavin depleted cells. Cells were seeded at 25,000/well with 2.5 ml of appropriate medium and incubated at 37 ºC for 160 hours. Incucyte software was used to image the wells in 16 segments every 2 hours.
Images **A-D** (navy blue box) show control cells grown in 1 µM riboflavin +10% FBS at 28, 76, 128 and 160 hours.
E-H (bright blue box) depict the 1 µM riboflavin +2% FBS grown controls at 28, 76, 128 and 160 hours respectively. Morphology and growth of 0.15 nM riboflavin-treated cells are shown by images **I-L** (pink box) at 28, 76, 128 and 160 hours. Arrows indicate elongated cells and a large group of dead cells in K are denoted by a round-ended arrow.

3.4. Differences in the mitochondrial proteome observed in riboflavin-restricted cells

SILAC was used to assess the changes in mitochondrial proteins between 143Bs grown in riboflavin-free DMEM + 2% FBS (0.15 nM riboflavin) and 143Bs grown in standard DMEM + 10% FBS. The experiment was carried out at two time points, informed by the OCR data collected previously – 5 days and 13 days in riboflavin-restricted media. Due to lack of availability of riboflavin-free SILAC media (to allow addition of heavy labelled arginine and lysine), see Chapter 2, section 2.7.1, only a single labelling orientation was possible. Although this data must be interpreted with caution - due to lack of repeats - it was informative, showing changes in a specific group of complex I subunits and a decrease in many mitoflavoproteins. SILAC data for riboflavin-deficient 143Bs after 5 days demonstrated lower protein coverage with ratios obtained for 1850 proteins compared to 2289 in the 13-day condition (Appendix 7.1). Complex I subunit coverage also varied: <50% were detected in the shorter treatment whereas ratios for 73% were measured in the longer incubation.

Figure 3.15 shows a similar spread of data points in both data sets: most proteins were unchanged in abundance, centring on Log_2 (fold change) = 0 (one-fold/no change), tailing-off to proteins with lower abundance in the riboflavin-depleted condition (positive H/L ratio) or increased levels in riboflavin-depletion (negative H/L ratio). There is some indication that the change in protein abundance was progressive, due to an increased number of proteins with higher or lower abundance after 13 days of riboflavin depletion compared to 5 days.

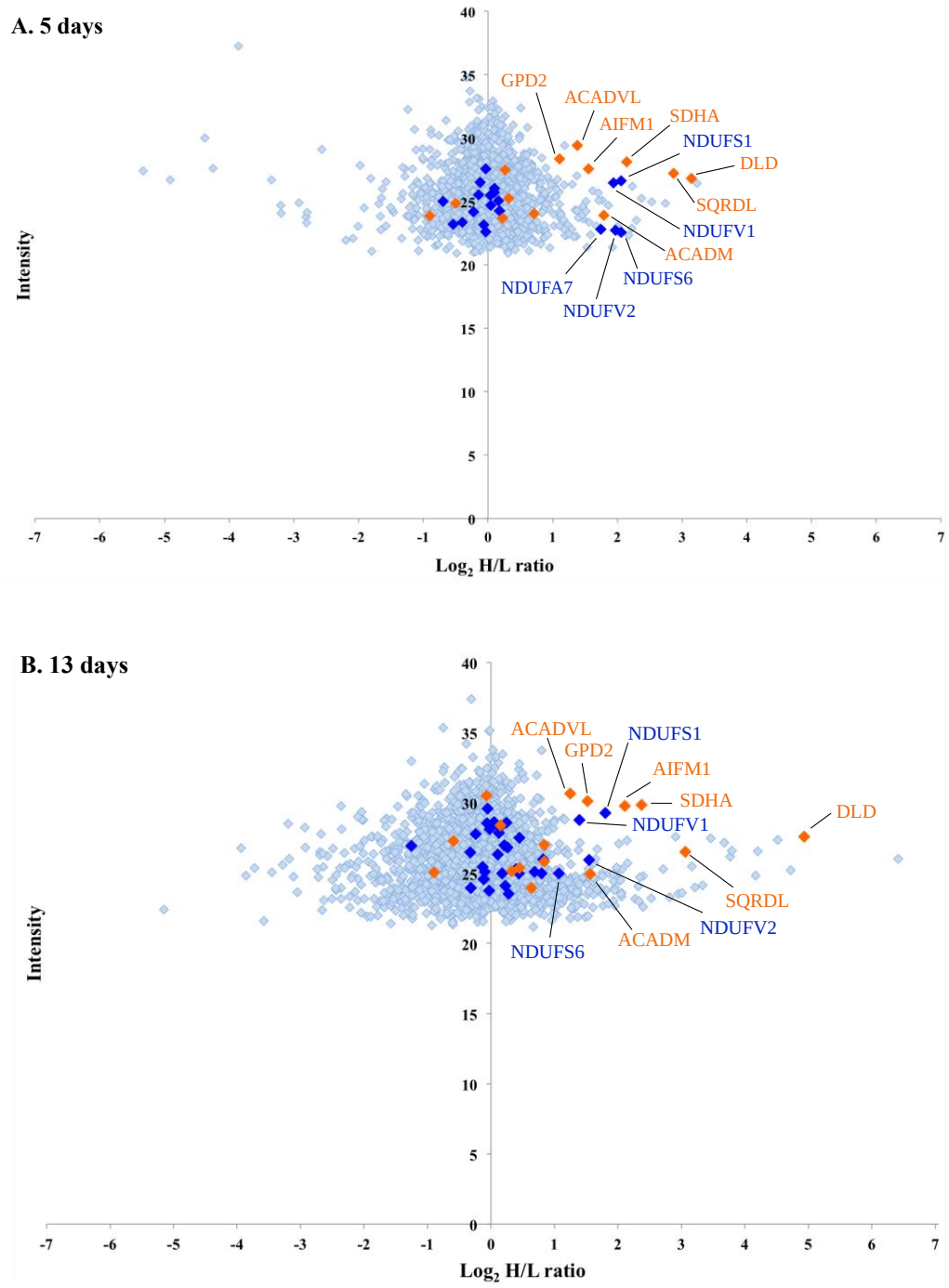


Figure 3.15 Single labelling orientation SILAC data for 143Bs grown in riboflavin-free DMEM + 2% FBS for 5 or 13 days. Data is displayed as Log_2 H/L ratio where positive H/L ratios represent proteins with higher abundance in the 143B control compared to the riboflavin-deficient condition. **A.** SILAC data for 5-day riboflavin-deficient incubation. **B.** SILAC data for 13-day riboflavin-deficient incubation. Complex I subunits are coloured dark blue and mitoflavoproteins shown in orange. From these two groups proteins with a greater than two-fold decrease in abundance in the riboflavin-deficient cells are labelled. Intensity refers to the total protein peak intensity (heavy and light labelled).

SILAC data presented in Figure 3.15 indicates growing 143B cells in 0.15 nM riboflavin-containing DMEM for short periods causes a change in protein abundance, with a particular subset of complex I subunits and mitoflavoproteins being decreased under these conditions. 20/45 complex I subunits were detected in the 5-day treatment and 32/45 in the 13-day treatment, and while most remained unchanged in abundance NDUFA7, NDUFS1, NDUFS6, NDUFV1 and NDUFV2 were the most changed subunits in each sample. These demonstrated above a two-fold decrease in abundance (NDUFA7 was just below this threshold in the 13-day treatment) and indicates that a subset of complex I subunits are detrimentally affected by the lack of flavin. Figure 3.16 shows the location of these subunits on the human model of complex I (generated as described in Chapter 6, section 6.6.) - highlighting that these proteins are structurally related. NDUFV1 is the flavin containing subunit and along with NDUFV2, NDUFS1 and NDUFS6 forms the NADH-binding (N-) domain, with four additional subunits. NDUFA7 is part of the Q-domain and sits just below the subunits of the N-domain. Therefore SILAC data suggest that subunits proximal to the bound cofactor FMN have lowered abundance when riboflavin is depleted and may indicate lowered stability of this region of the enzyme. However the SILAC only measures the overall abundance change and so the proportion incorporated into the overall enzyme cannot be determined.

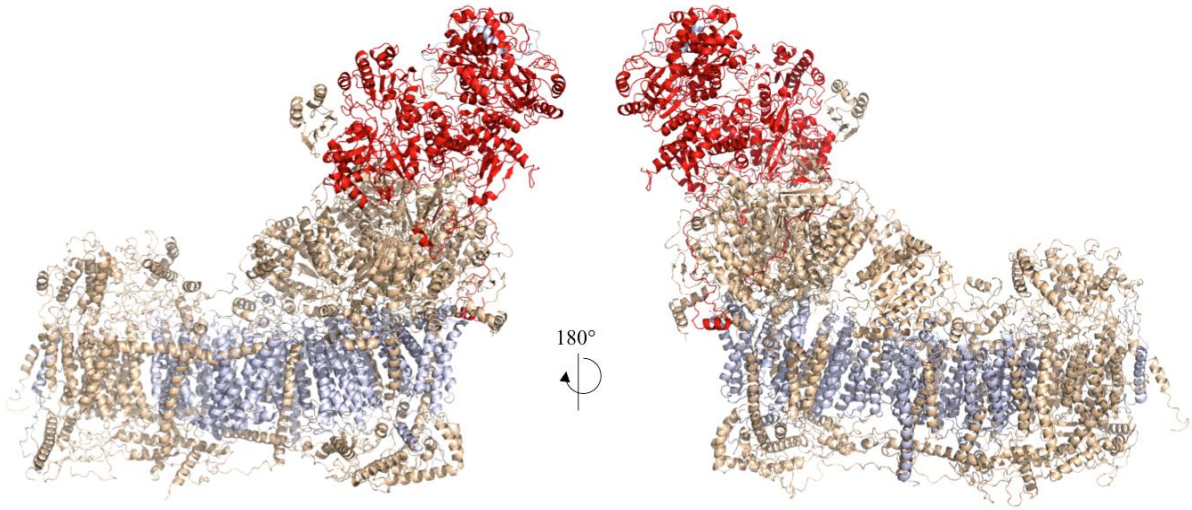


Figure 3.16 Structural location of subunits showing a greater than two-fold decrease in abundance after riboflavin deletion to 0.15 nM in 143B cells. The human model of complex I (Chapter 6, section 6.6) is shown in wheat with the five subunits NDUFA7 NDUFS1, NDUFS6, NDUFV1 and NDUFV2 are coloured in red – showing their position at the top of the hydrophilic arm surrounding the bound flavin factor FMN. Subunits that were not detected in either SILAC experiment are shown in grey.

Additionally 13 mitochondrial flavoproteins were detected in the 5-day treatment and 16 in the 13-day treatment sample, but only a small proportion of these were considerably changed in abundance (seven in the 5-day treatment and eight in the 13-day treatment - including NDUFV1). Figure 3.16 highlights these proteins and comparison between the two treatment durations shows the same subset was affected. Notably these include the FAD-containing subunit of complex II (SDHA), which suggests further disruption to the respiratory chain and may contribute to the altered OCR observed in low riboflavin

conditions, Figure 3.11. The two mitoflavoproteins showing the greatest decrease were dihydrolipoamide dehydrogenase (DLD) and sulphide:quinone oxidoreductase (SQRDL), which were more than five-fold less abundant in cells grown in 0.15 nM riboflavin compared to riboflavin-rich conditions. DLD is an E3 component of dehydrogenase complexes and reduces NAD^+ and SQRDL is involved in sulphur metabolism and reduces ubiquinone. Therefore they both influence the redox status of reducing equivalents within the mitochondria and so alterations in the abundance of these proteins may indicate wider affects on mitochondrial function. Despite these cells experiencing a high level of stress and not remaining viable for an extended period these data suggest that restricting riboflavin availability affects complex I and has wider effects on the mitochondria. This was then followed up in more detail with experiments undertaken on the conditioned RFK knockdown 143Bs (section 3.6.).

3.5. Learning outcomes and experimental development

From this work different approaches to limiting riboflavin availability had varying effects on 143B cells:

- treatment with lumiflavine on a background of standard growth conditions had no observable impact.
- decreasing riboflavin in the medium by riboflavin deficient DMEM substitution (to generate conditioned cells after one month) provided a basis for further treatment resulting in a respiratory phenotype:
    - severely lowered FBS supplementation (2%) resulted in a respiratory phenotype at 0.15 nM riboflavin and no long term viability.
    - lumiflavine treatment of conditioned cells (with 10% FBS) increased cell stress and resulted in a low basal OCR, with no long term viability - showing how severe riboflavin deficiency impacts cells survival.

These learning outcomes informed the development of future experiments using cultured mammalian cell lines. They showed that extracellular riboflavin was highly influential when manipulating flavin handling; and therefore ‘conditioned’ cells (grown in riboflavin deficient DMEM + 10% FBS for one month) were used in further investigations. Consequently, due to the short-lived and transient nature of siRNA treatment it was necessary to find an alternative technique to target protein expression for long periods, to allow consistent study of the effects under riboflavin-depleted conditions. Therefore a comparable approach was chosen: stable knockdown by shRNA.

3.6. Stable knockdown of flavin handling proteins in 143B cells to allow thorough investigation of riboflavin depletion

3.6.1. Stable knockdown approach to target cellular flavin processing

The stable knockdown of proteins by shRNA is a widely used tool for manipulation of animal cells. In contrast to siRNA-mediated knockdown, shRNA is synthesised by the host cells after delivery by a DNA vector, which is introduced by viral transfection. Exogenous RNA oligonucleotides, such as siRNA, that are introduced into the cell are degraded, which allows the target mRNA levels to recover and protein expression to return to normal. In contrast using shRNA allows the knockdown to be maintained and the DNA plasmid encoding the specific shRNA can integrate into the host genome, which is achieved by utilising the lentivirus system.

As described in Chapter 2, section 2.3.2. the second-generation lentiviral system was chosen along with the pLKO.1-TRC cloning vector (referred to as pLKO.1), for shRNA insertion. This system uses separate plasmids for the packaging and envelope proteins (psPAX2 and pMD2.G respectively) required to encapsulate the shRNA-encoding vector. This replication incompetent system splits up the components that form an infective virus particle and so overcomes the risk of self-replication. Three plasmids were transfected into HEK 293T cells using the Fugene® agent. The lentviral particles produced were then harvested from the growth medium and used to infect the host 143B cell line with the shRNA-containing plasmid.

3.6.1.1. shRNA knockdown of flavin-manipulating enzymes

RFK and SLC25A32 were chosen for stable protein knockdowns in 143Bs, in addition to FADS1 due to its role in FAD synthesis from FMN.

pLKO.1 vectors containing each of the shRNAs were purchased from Sigma. Sigma MISSION® shRNA bacterial glycerol stocks were used, with two shRNA constructs for each protein. Unfortunately, despite successful production of lentiviral particles by HEK293T cells and subsequent infection of 143B cells (described in Chapter 2, section 2.3.2.), which yielded puromycin resistant colonies and suggested successful introduction of the vector, screening by Western Blot only revealed successful knockdown of RFK. The shRNA vector TRCN0000037602 yielded puromycin resistant 143Bs displaying significantly decreased RFK expression (RFK KD), validated by quantitative Western Blot to show a reduction of approximately 90% (Figure 3.17).

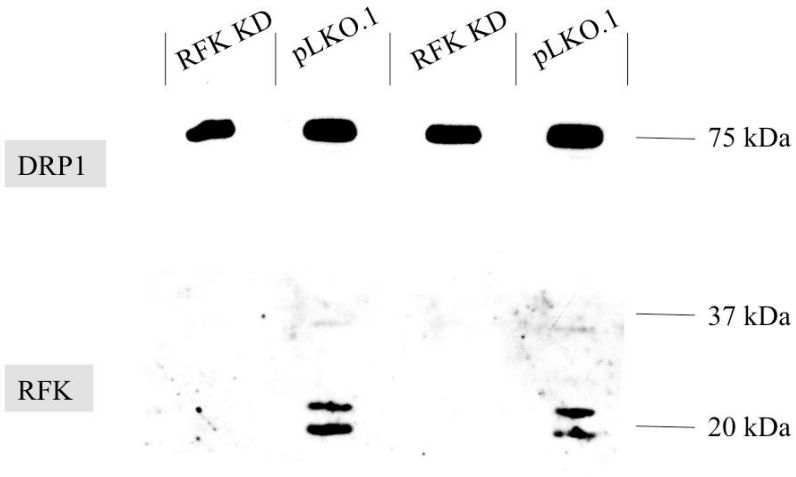


Figure 3.17 Riboflavin kinase knockdown by shRNA shown in Western Blot analysis of cell lysates. Five days after infection with lentiviral particles a sample of shRNA-expressing 143Bs were harvested and cell lysate prepared. 20 µg protein was loaded onto a NuPAGE™ 4-12% Bis-Tris gel for separation by SDS-PAGE. Proteins were then transferred onto PVDF membranes and blotted for RFK (expected band at 17.6 kDa, antibody from Proteintech) and DRP1 (as a loading control, expected band at 75 kDa, Abnova). Vector expressing 143Bs (pLKO.1) demonstrated RFK expression, seen as a double band ~20 kDa, whereas the stable RFK knockdown from TRCN0000037602 vector expression showed no observable bands by ECL. This was repeated and then quantified by fluorescently coupled antibodies to obtain a knockdown efficiency measurement of 88%.

For all future experiments a vector control was used, this was a 143B CyWT line expressing pLKO.1 that was puromycin resistant, but did not express an shRNA.

Due to the low success rate of the Sigma MISSION® shRNA plasmids a number of shRNA were designed, across the whole transcripts for SLC25A32 and FADS1, using the online tool from the Whitehead Institute (Yuan *et al*., 2004). Three shRNA sequences were chosen for each target, which were then inserted into the pLKO.1 vector - after optimisation of the cloning technique, as discussed in Chapter 2, section 2.3.2. However again there was no observable knockdown in the target protein, therefore the RFK KD was the only stable knockdown in 143Bs taken forward for characterisation.

3.7. Investigating the phenotype of the stable riboflavin kinase knockdown, using conditioned 143Bs

In order to characterise the RFK knockdown, 143B cells expressing the shRNA were gradually introduced to riboflavin-free DMEM (containing 1 µg/ml puromycin), alongside the pLKO.1 control. Their morphology and growth profile were followed by image capture and cell density tracking using Incucyte software, which demonstrated that conditioned RFK knockdown cells did not display altered growth rate or cellular morphology and were viable after sustained passaging.

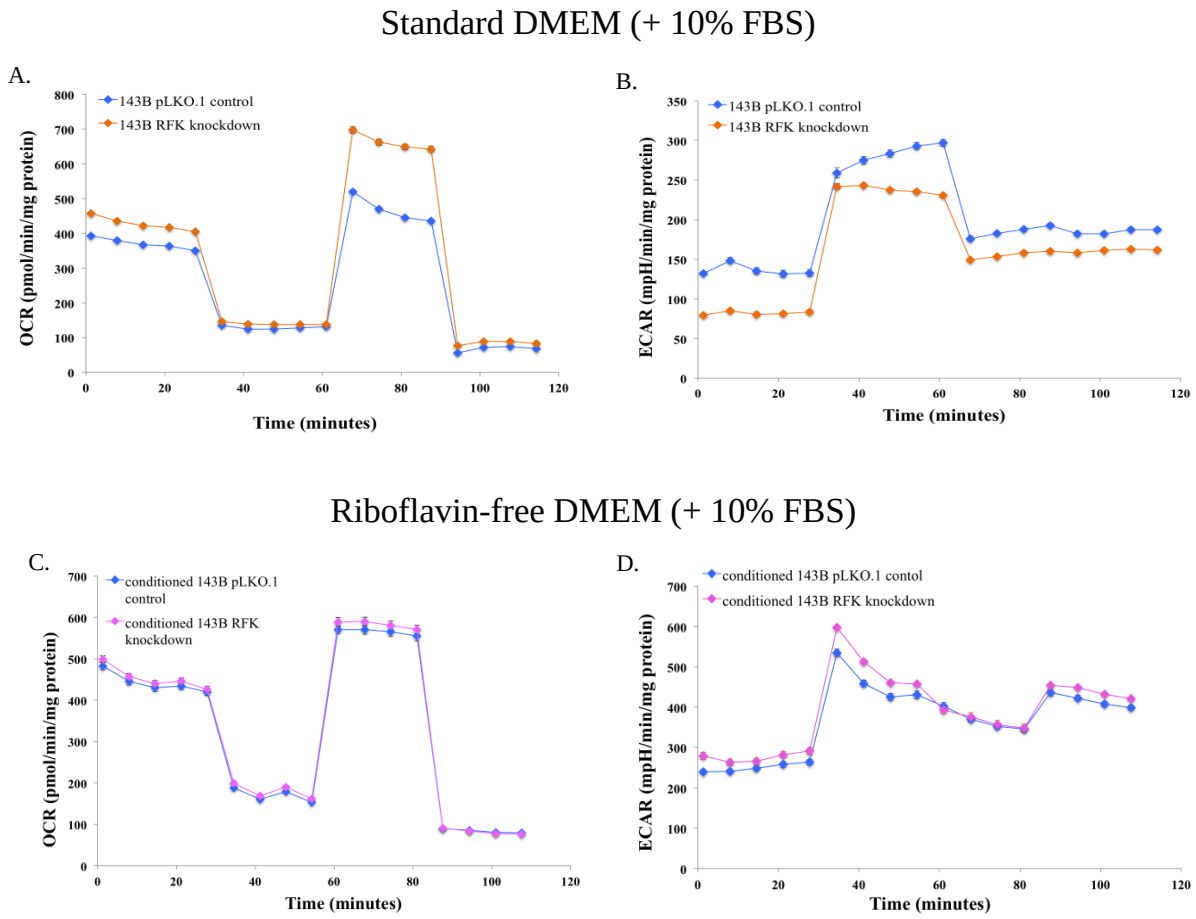


Figure 3.18 Seahorse extracellular flux data comparing the respiratory profile of stable RFK knockdown in 143Bs to a vector expressing control. A. Seahorse extracellular flux data comparing the OCR of control pLKO.1 143Bs to the RFK stable knockdown, grown under standard conditions. Cells were seeded at 15,000 well the day before the assay. 46 wells of each condition were assayed and the experiment repeated thrice (mean results for one representative assay shown, with error bars denoting standard error). **B.** Corresponding ECAR data for 143Bs grown in 1 µM riboflavin-containing DMEM. **C.** Seahorse extracellular flux data comparing the OCR of control pLKO.1 143Bs to the RFK stable knockdown, grown in riboflavin-depleted conditions. Cells were seeded at 15,000 well the day before the assay. 46 wells of each condition were assayed and the experiment was repeated thrice at different time intervals after one month conditioning, to track any respiratory phenotype changes (mean results from the earliest assay are shown, with error bars denoting standard error). The respiratory profile was representative of that obtained in all repeats. **D.** Corresponding ECAR data for 143Bs grown in riboflavin-depleted conditions.

RFK knockdown cells grown under standard conditions, with abundant riboflavin, were assessed by a stress test to compare the OCR to the control (Figure 3.18A&B). Once conditioned, RFK knockdown and control cells were assessed again in the same way (Figure 3.18C&D). The respiratory profile was unaffected by RFK knockdown in 143Bs and even with the additional strain of riboflavin depletion the OCR was equivalent to the control (Figure 3.18A&C), indicating that OXPHOS capacity was retained and RFK knockdown had a low impact on FMN incorporation into complex I. Basal respiration was unaffected and addition of oligomycin and FCCP had an equivalent effect to the control; therefore spare respiratory capacity was also unchanged. The extracellular acidification rate also demonstrated no variation between the RFK knockdown and control cell lines (Figure 3.18B and D). A difference in basal ECAR may be an indication of a shift towards a more glycolytic phenotype, and would suggest impairment in OXPHOS; however this was not observed, despite additional stress introduced by conditioning.

To further probe the impact of RFK knockdown on complex I activity, crude mitochondria were prepared and complex I in-gel activity assays undertaken. Figure 3.19 shows the position of monomeric complex I from cell lines was equivalent to that of bovine membranes; an additional higher band is often observed in cell line samples and may suggest the presence of two complex I populations. A band corresponding to complex V was also visible, demonstrating that there was slightly more protein loaded for the RFK KD sample compared to the control, which ameliorates the difference in active complex I band intensity observed. This suggested that decreased RFK expression did not affect complex I activity substantially.

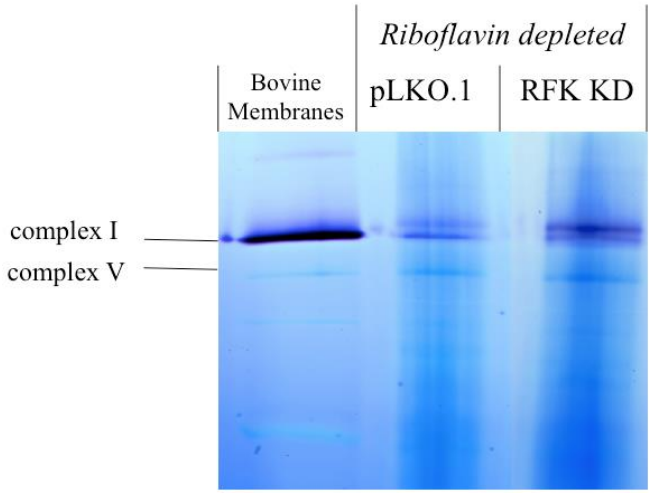


Figure 3.19 BN-PAGE with complex I in-gel activity stain for riboflavin depleted 143B cells. Digitonised cells were prepared from conditioned RFK knockdown and pLKO.1 control 143Bs and after DDM solubilisation ~10 µg was loaded onto a NativePAGE™ 3-12% Bis-Tris gel, along with DDM solubilised bovine mitochondrial membranes as a standard. Complex I activity was assessed using the electron acceptor NBT. 0.5 mg/ml NBT and 100 µM NADH in 20 mM Tris pH 7.4 were added to the gel after washing in cold dH_20 to remove excess Coomassie stain and retain activity. After 30 minutes incubation the reagent mixture was removed and the gel washed in dH_20. Complex I was stained purple if flavin activity was present, labelled band.

3.8. Complexome analysis of riboflavin kinase knockdown cells under riboflavin restricted conditions

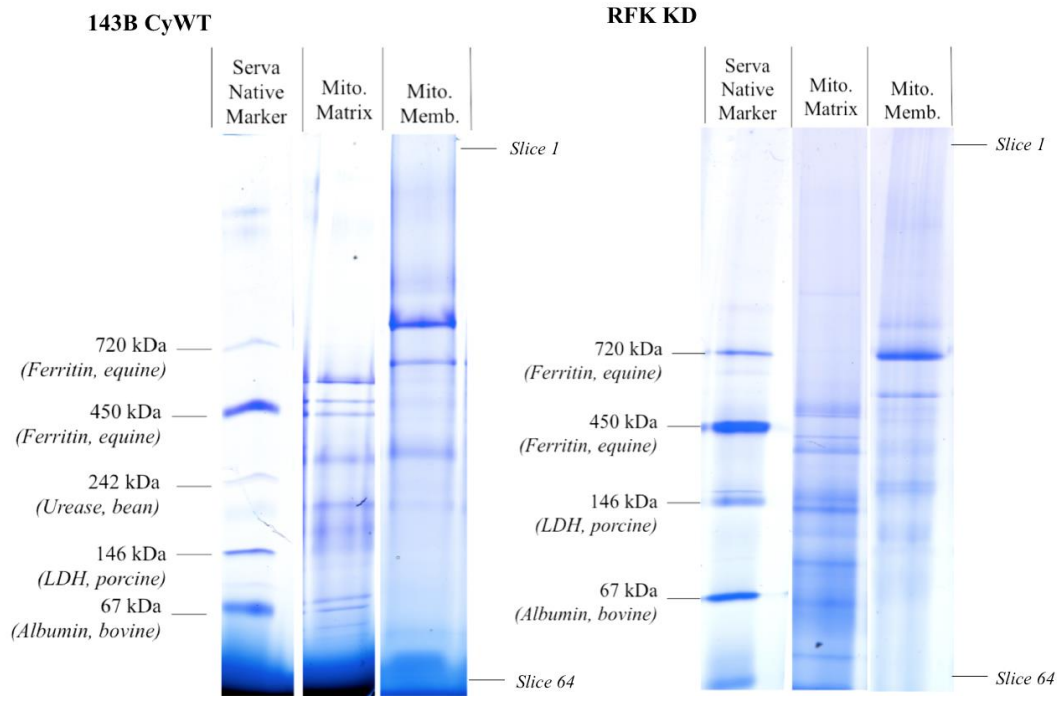


Figure 3.20 Mitochondrial matrix and membrane fractions from 143B CyWT and conditioned RFK KD 143Bs. Mitochondria were purified from 143B CyWT and RFK KD cells, as described in Chapter 2, section 2.4.2. Matrix and membranes were then separated by sonication. Mitochondrial membranes were solubilised in 1.7% DDM prior to BN-PAGE. Different amounts of membranes and matrix were loaded to obtain the optimal protein loading for analysis by mass spectrometry. Shown above are the two lanes chosen for 143B CyWT (left panel, prepared by Dr Hannah Bridges) and RFK KD – containing 10 µg protein (right panel). The Serva Native ladder was used to estimate protein masses. Once run the gel was stained with Coomassie before submission for complexome analysis - both the mitochondrial membranes and matrix lanes were cut into 64 even slices, direction of slice numbering shown.

Figure 3.20 displays the mitochondrial membrane and matrix fractions from control 143B CyWT and conditioned RFK knockdown 143Bs. The samples were prepared as described in Chapter 2, section 2.7.2., cut into 64 1mm slices and analysed by mass spectrometry to determine the distribution of proteins and possible associations between them. Therefore the complexomes obtained referred to the matrix and mitochondrial membrane fractions separately. It was hoped that this would provide a more detailed understanding of any aberrations observed.

The 143B CyWT samples were prepared and run previously by Dr Hannah Bridges, with cells cultured in standard DMEM + 10% FBS. These data were compared with the conditioned RFK knockdown complexomes to ascertain whether knocking down RFK and limiting flavin availability had an observable impact on the respiratory complexes. This could then be extended to use a specific riboflavin depleted/conditioned 143B control, if sufficient differences were observed. In order to compare the two

data sets protein mass calibration curves were generated, therefore results are described with relation to the estimated mass of the protein/complex from each gel (rather than the slice number). Calibration curves for the membrane data sets are shown in Figure 3.21 and proteins used given in Appendix 7.2A&B. Data for the ETC complexes from both membrane samples is given in Appendix 7.2C&D.

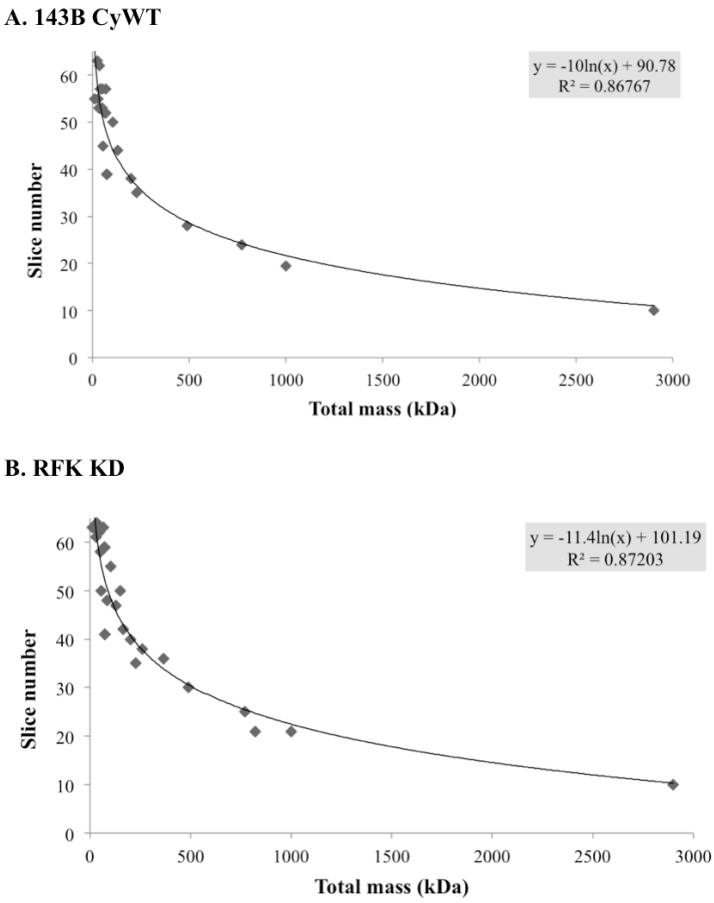


Figure 3.21 Protein mass calibration curves for membrane fraction complexomes. To allow comparison between the two complexome data sets proteins or complexes of known molecular weight were chosen and their position in the gel correlated to their mass. More than 20 mitochondrial proteins/complexes from each complexome were chosen. Despite being mitochondrial membrane fractions there is known to be some contamination with soluble mitochondrial elements and other cellular compartments, therefore these proteins were useful for mass calibration, along with mitochondrial transporters and the respiratory complexes themselves. **A.** 143B CyWT. **B.** RFK KD. Protein masses were obtained from the UniProt database (UniProt Consortium T, 2018). Table of proteins used in Appendix 7.2A&B.

3.8.1. Mitochondrial membrane complexome analyses of conditioned RFK knockdown cells

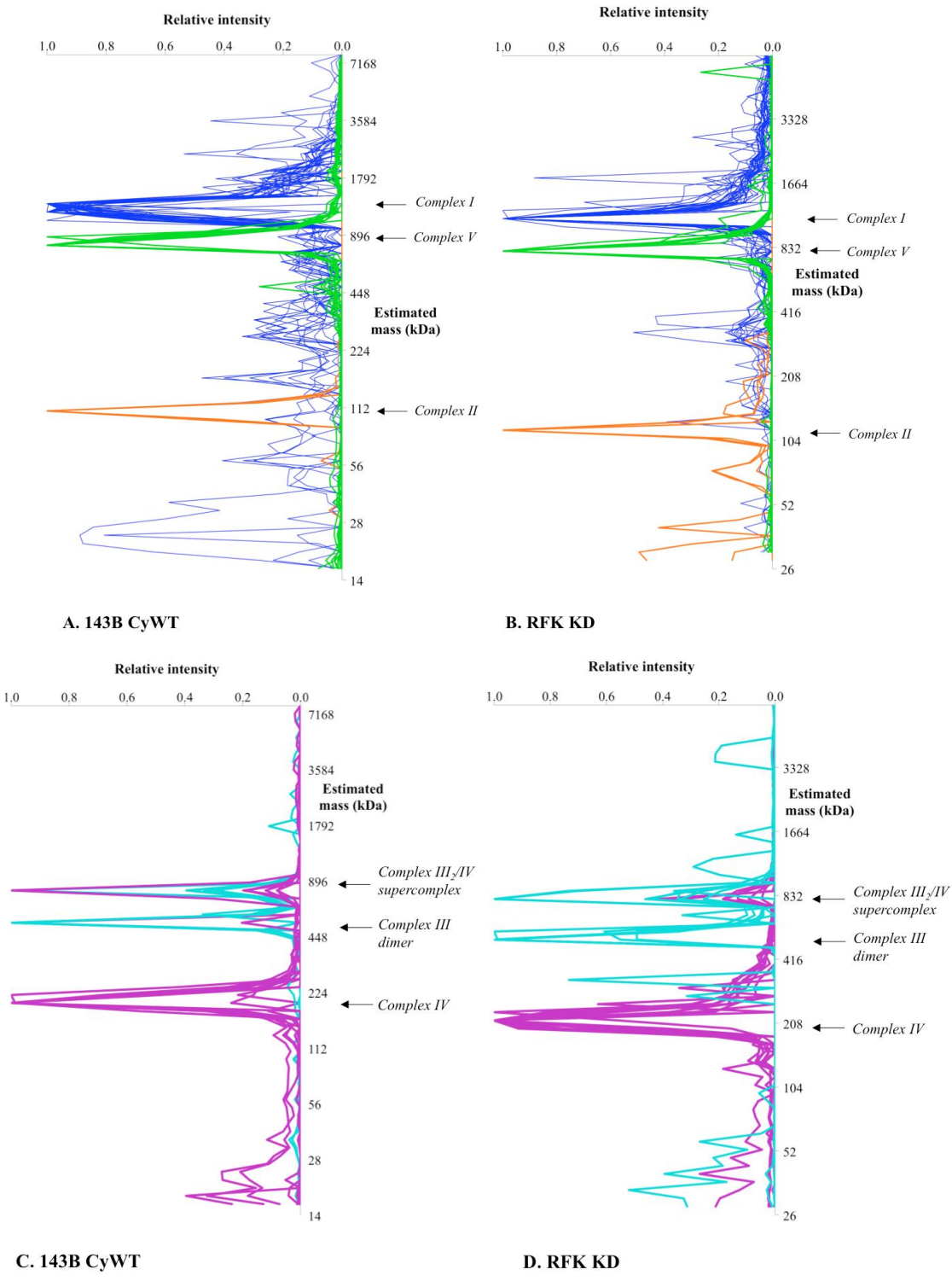


Figure 3.22 Distribution of respiratory complexes across gel slices after BN-PAGE. Solubilised mitochondrial membranes were separated by BN-PAGE, to maintain native complexes. Coomassie stain was used before submission for complexome analysis - both lanes were cut into 64 even slices. **A.** 143B Cybrid WT control membrane - positions of complex I (dark blue), II (orange) and V (green). **B.** Conditioned RFK knockdown 143B membrane complexome positions of complex I, II and V. **C.** 143B Cybrid WT control membrane – positions of complex III (cyan) and complex IV (purple). **D.** RFK knockdown 143B membrane complexome – positions of complex III and complex IV. Complexes are grouped into different panes for clarity.

Figure 3.22 shows the distribution of each respiratory complex across the two membrane samples. Using the calibration curves, complex I (calculated mass ~ 1000 kDa) was estimated to migrate at 1249 kDa in the control and the highest intensities for the detected subunits spanned two neighbouring slices. Comparatively, complex I in the RFK KD membrane sample migrated at approximately 1135 kDa and the subunits peaked in one gel slice. 38/45 complex I subunits were detected in 143B CyWT (lacking NDUFA2, NDUFB2, NDUFC1, NDUFAB1 (x2), ND4 and ND4L) and 41/45 in RFK KD (lacking NDUFA1, NDUFC1, ND3 and ND4L). The core mitochondrially encoded-membrane domain subunits are often the most difficult to detect by mass spectrometry due to their hydrophobic nature and low number of tryptic peptides. Additionally very small subunits, such as NDUFC1, are also less commonly measured. However, there was good coverage of the subunits across both data sets. A shoulder was present on the higher molecular weight side of the main complex I peak in each sample that may indicate the present of dimerised complex I. In general there were no substantial differences between the complex I structure and migration pattern across the two complexomes.

In both complexomes all complex II subunits were detected (SDHA, B, C and D); migrating at 108 kDa in the control and 116 kDa in RFK knockdown (calculated mass 130 kDa). Figure 3.22A and B show that in both profiles the clustering of the subunits was consistent, with a clean single peak around the expected mass. There was no indication of misassembled complex II despite restricted flavin availability in the conditioned RFK knockdown.

Of the 11 complex III (cytochrome b-c1 complex) subunits both membrane samples showed good coverage (9/11). RFK knockdown complex III migrated to the expected mass of 515 kDa (calculated mass of dimer 490 kDa), whereas the control complex was found at 533 kDa. Figure 3.22C and D demonstrate that two peaks were observed; in the 143B CyWT sample this second clustering at 795 kDa contained eight subunits, all with low relative intensities except subunit 6 (highest intensity peak); representing the III_2/IV supercomplex (calculated mass 690 kDa); supported by the presence of most complex III subunits and some complex IV subunits (including a high intensity of subunit 7A2L). The RFK knockdown mirrored this dual peak profile, with the higher mass peak nearer to the expected III_2/IV mass (estimated mass 799 kDa).

The complexome profile of cytochrome c oxidase (complex IV) was generally consistent between the two samples, Figure 3.22C and D. A main peak of subunit intensities was observed for the complex (calculated mass 200 kDa) and another higher molecular weight peak coincided with the second peak in complex III subunits, supporting the presence of the III_2/IV supercomplex. 12 subunits were detected in 143B CyWT and 14 subunits present in the RFK knockdown. Both samples showed a high intensity peak near the expected monomeric complex mass (calculated 200 kDa: 196 kDa in 143B CyWT and 214 kDa in RFK knockdown).

The final complex of the electron transport chain, ATP synthase, was visible on BN-PAGE after staining and is often used as a marker for mitochondrial quality/purity, due to its abundance and consistent migration. In both samples 13 subunits were identified, 143B CyWT showed good clustering and a clean peak at 795 kDa (calculated mass 770 kDa), shown in Figure 3.22A. The subunit coupling factor 6 was found to peak in the slice above the rest of the subunits, at 859 kDa, but peaked with the complex in the knockdown. The RFK knockdown sample peaked at 799 kDa and also displayed a clean profile, Figure 3.18B. In the control sample there was also a noticeable secondary peak, albeit of low subunit intensities, at 482 kDa that may indicate a subassembly; this was also present in the RFK knockdown (432 kDa). Both samples additionally contained the inhibitory protein IF1, with its peak intensity neighbouring that of ATP synthase. This agreement between migration patterns of both subunits and IF1 indicated that ATP synthase was unaffected by cellular changes resulting from RFK knockdown and riboflavin depletion and was supported by other experiments showing normal respiration by these cells.

Overall, by comparison of the two complexomes, it was found that all respiratory complexes were largely unaffected by the knockdown, with small changes likely to be attributed to inherent differences in sample preparation or BN-PAGE analyses.

3.8.2. Complex I assembly factor migration patterns were unaffected in conditioned RFK knockdown cells

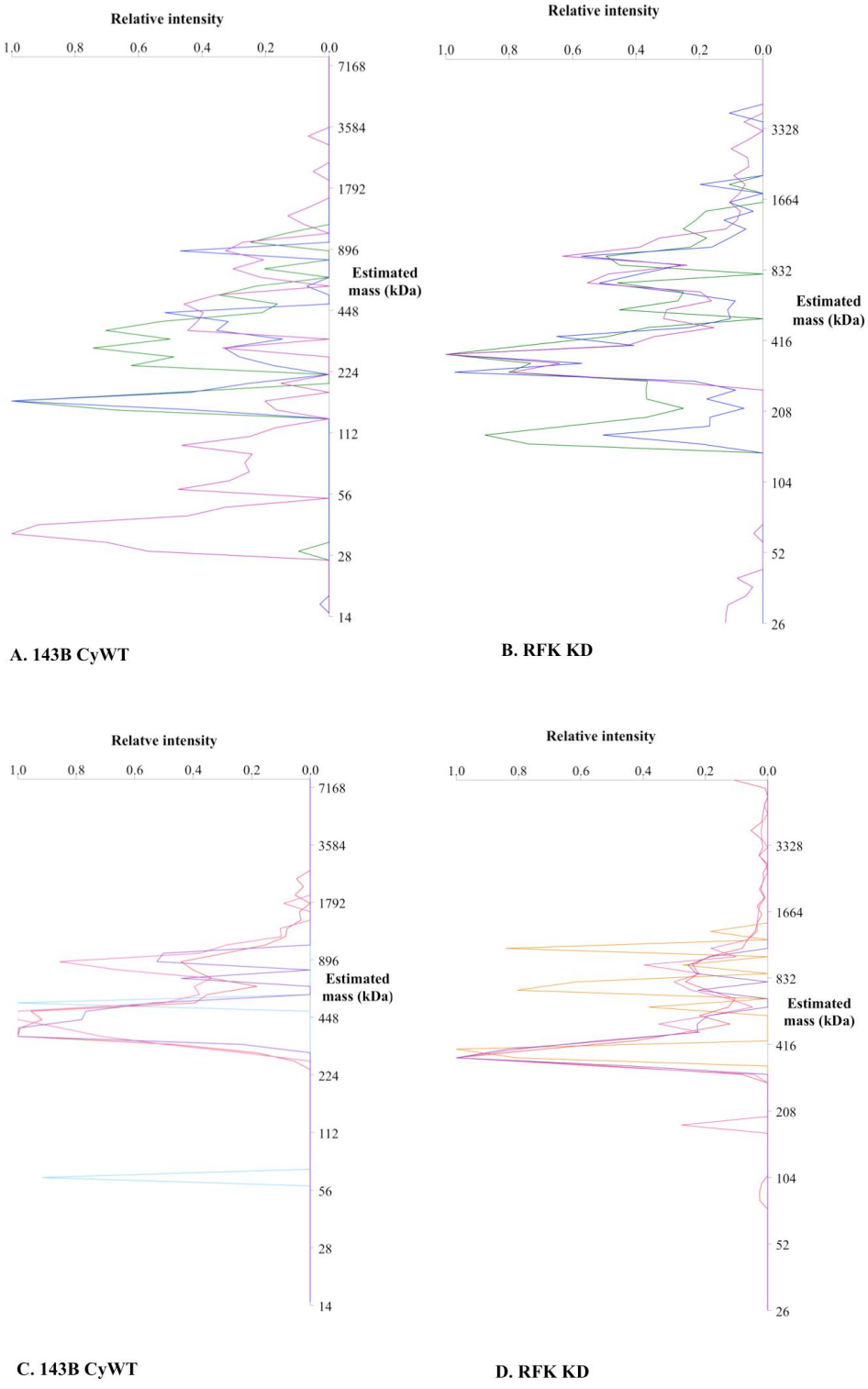


Figure 3.23 Clustering of two groups of complex I assembly factor across gel slices after BN-PAGE. A. 143B CyWT control membrane complexome showing NDUFAF3 profile in green, NDUFAF4 in blue and TIMMDC1 in pink. **B.** Conditioned RFK knockdown 143B membrane complexome - NDUFAF3, NDUFAF4 and TIMMDC1. **C.** 143B Cybrid WT control membrane complexome showing ACAD9 profile in red, NDUFAF1 in purple, COA1 in light blue and ECSIT in pink. **D.** Conditioned RFK knockdown 143B membrane complexome - TMEM126B (orange,) ACAD9 (red), NDUFAF1 (purple) and ECSIT (pink).

From the complexomes 10 assembly factors were detected in both 143B CyWT and RFK knockdown membrane samples. Two main groups were present, demonstrating assembly factors known to cooperate in assembling specific complex I subunits. The earliest stage assembly factors are NDUFAF3 and NDUFAF4, found bound to the growing Q-domain and thought to facilitate attachment to the P_P-a subassembly (Figure 3.24, bright green) of the membrane arm, along with TIMMDC1. In Figure 3.23A, 143B CyWT shows clustering of NDUFAF3 and NDUFAF4 as expected with a peak at 161 kDa, which also contains subunits NDUFA5, NDUFS2, NDUFS3, NDUFS7 and NDUFS8 at low intensities. This correlates well with an early subcomplex with calculated mass of 170 kDa. However TIMMDC1 does not comigrate with its associated assembly factors and is only found at low molecular weights, peaking at 36 kDa, suggesting its presence as discreet monomer (32 kDa). In contrast the RFK knockdown complexome shows all three assembly factors comigrating and peaking together at 363 kDa, Figure 3.23B. Along with the presence of subunits NDUFS2, NDUFS3, NDUFS7 and NDUFS8 this indicates the presence of the subcomplex comprising the core Q-domain, three assembly factors and ND1 (calculated mass of 237 kDa). TIMMDC1 was also found at low intensities at its estimated monomeric mass. NDUFAF3 and NDUFAF4 were also found in high intensities at a lower molecular mass of 165 kDa and could be the earliest assembly factor bound Q-domain, with NDUFS2, NDUFS3, NDUFS7 and NDUFS8 all present in this region.

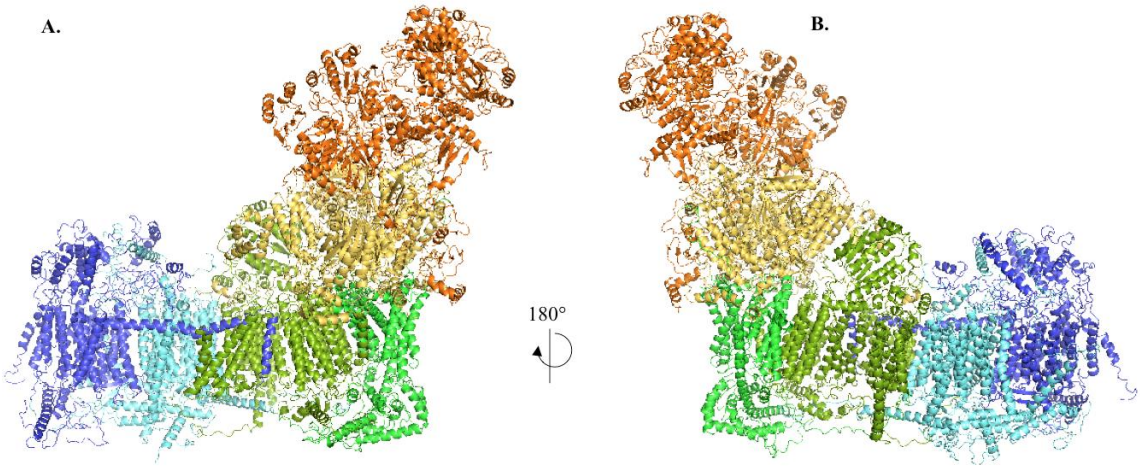


Figure 3.24 Complex I structure coloured by domains. The most recent assembly pathway describes the construction of the enzyme split into structural domains: N-domain (orange), Q-domain (yellow), peripheral membrane arm (P_P greens) and distal membrane arm (P_D blues). This is further divided into two subunit collections for each membrane arm region: peripheral into P_P-a (bright green) and P_P-b (dark green), distal into P_D-a (pale blue) and P_D-b (dark blue) due to their assembly into discreet subcomplexes. Naming reflects that of Nijtmans and co-workers (Guerrero-Castillo *et al.*, 2017). The human model of complex I was constructed using mouse complex I cryo-EM structures (Chapter 6, section 6.6).

Another group of assembly factors, known to cooperate in the assembly of the membrane arm P_P-b domain (Figure 3.24, dark green), are ACAD9, ECSIT, NDUFAF1, TMEM126B, TMEM186 and COA1. In both complexomes ACAD9, ECSIT, NDUFAF1 were detected, along with COA1 in 143B CyWT and TMEM126B in the RFK knockdown. Figure 3.23C shows that in the 143B CyWT data set COA1, NDUFAF1, ACAD9 and ECSIT displayed a similar migration but their peak intensities differ across 358-533 kDa, with only NDUFAF1 and ACAD9 peaking together. COA1 is also present at high intensity at a lower molecular mass (65 kDa), which may be due to its role in cytochrome c oxidase assembly. In contrast the RFK knockdown data for these assembly factors shows much clearer cooperation, in Figure 3.24D, with highest intensities for all four (TMEM126B present but COA1 not detected) across two neighbouring slices at 363-396 kDa, which may indicate the early P_P-b subcomplex. The remaining assembly factors detected, NDUFAF2, FOXRED1 and TMEM70, displayed different migration patterns, which did not correspond well to their proposed functions, therefore no valuable interpretation could be made from their profiles; it is possible that the strength of interactions for these factors is lower than those displaying a more consistent clustering pattern.

Overall the assembly factors detected in the RFK knockdown membrane complexome displayed a consistent pattern with that proposed by Guerrero-Castillo *et al.*, particularly the previously known cooperatively of NDUFAF1, ACAD9 and ECSIT and between NDUFAF3 ad NDUFAF4 at the early stages of assembly. Despite differences in the profiles across the two samples it can be concluded that conditioned RFK knockdown cells did not adversely affect complex I assembly in 143Bs, as many assembly factors were detected, most of which displayed migration patterns and associations matching those from previously published research.

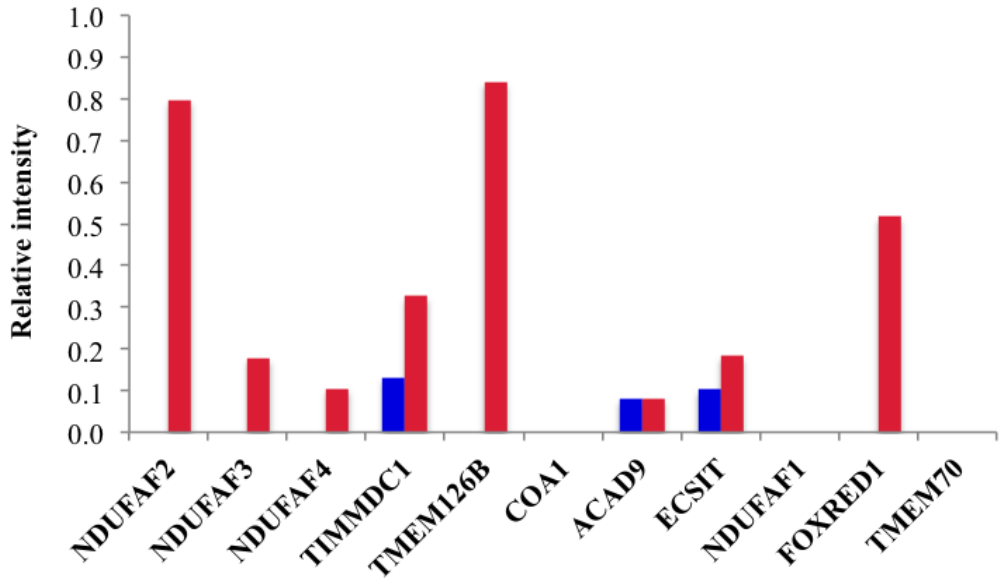


Figure 3.25 Complex I assembly factors present with mature complex I in membrane complexome samples slices. More complex I assembly factors were found to be present in the same slice as mature complex I in the RFK KD sample (red bars) compared to the 143B CyWT control (blue bars). COA1 was not detected in the RFK KD and TMEM126B was not detected in the 143B CyWT.

Interestingly assembly factors were found associated with the mature complex in RFK knockdown and not in the 143B CyWT. The point at which assembly factors leave the nascent enzyme has long been discussed and there is general agreement that on formation of the complete complex, with addition of the N-domain, the assembly factors all dissociate releasing the enzyme into a catalytically capable state. Yet in RFK KD 8/10 factors are still present, with NDUFAF2 and TMEM126B at a high relative intensity, Figure 3.25. It could therefore be hypothesised that the lowered availability of flavin may have resulted in disruption of the N-domain assembly – FMN incorporation into NDUFV1 – and as a result the assembly factors have not been released from the completed enzyme. Despite this, conditioned RFK knockdown cells have shown little differences compared to the control; section 3.7 demonstrated that RFK knockdown did not show altered oxygen consumption or responses to drug additions during a stress test and in-gel activity assay of complex I was equivalent to the control.

3.8.3. Complex I assembly intermediates were present in conditioned RFK knockdown cells

Assembly of complex I involves a series of subcomplexes, that are sequentially built together to form the mature, active enzyme. The assembly pathway published by the Nijtmans group describes nine named subcomplexes and an additional eight smaller assemblies. The possible subcomplexes found in the two complexomes (Tables 3.3, 3.4) were compared to this pathway and their identities inferred from subunit composition and estimated mass.

The subassemblies were detected over a range of gel slices, making their identification less definitive but from each complexome a number of subassemblies were observed that contained subunits and assembly factors consistent with those reported in assembly pathway studies. The complex I profile from the 143B CyWT membrane complexome (Figure 3.26A) displayed three possible lower molecular weight clusters of subunits. The largest of these (Table 3.1, subcomplex 1: 650-719 kDa) could be a late stage assembly intermediate as the subunits found in this region span the entire complex. A small 292-357 kDa subcomplex had three possible identities as shown in Table 3.1 (subcomplex 2); some subunits from each intermediate were present and the total masses were similar: Q/P_P-a included assembly factors NDUFAF3, NDUFAF4 and TIMMDC1, with P_D-a and P_D-b suggested by the presence of hydrophobic subunits known to associate as discreet intermediates. The smallest subcomplex (subcomplex 3: 161 kDa) most closely resembles the core Q-domain with NDUFAF3 and NDUFAF4 bound.

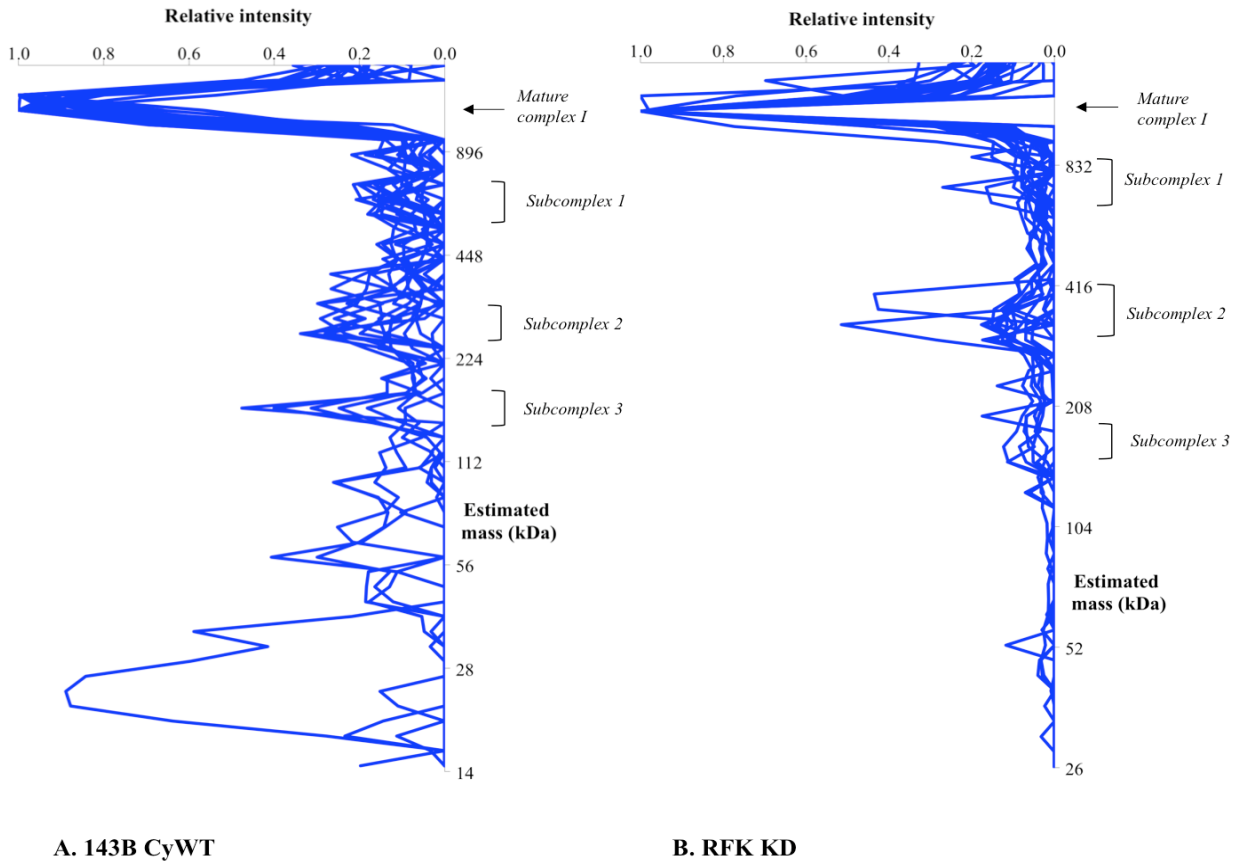


Figure 3.26 Clustering of complex I subunits at lower molecular weights indicate the present of subcomplexes. A. 143B Cybrid WT control membrane complexome showing the lower end of the complexome gel, from mature complex I. Subunits that displayed some relative intensity at lower molecular weight are shown in blue and possible subcomplexes highlighted. **B.** Conditioned RFK KD 143B membrane complexome - showing the lower end of the complexome gel, from mature complex I, with possible subcomplexes highlighted.

Figure 3.26B also displays a number of smaller complex I subunit clusters, indicating the presence of assembly intermediates in the RFK knockdown membrane fraction. These are shown in Table 3.2 with subunits found in the larger peak of 664-724 kDa spanning most of the complex and include seven assembly factors, which may be a mix of P_P-b/P_D-a (680 kDa) and Q-P_P (736 kDa) intermediates. Therefore large late stage subcomplexes were observed in both membrane fractions. Subcomplex 2 in the RFK knockdown was difficult to assign, with three possible identities due to the proteins present (Table 3.2). At its largest this collection could comprise an early P_P-b subcomplex (292 kDa) of mostly assembly factors or Q/P_P-a subcomplex (237 kDa) with NDUFAF3, NDUFAF4 and TIMMDC1 present. The other possibilities are less reliable, with fewer subunits present and lower total mass – P_D-a (230 kDa) or P_D-b (154 kDa). As in the control, the smallest intermediate (subcomplex 3: 165 kDa) resembles the core Q-domain. Despite the lack of definitive assignment the presence of smaller subcomplexes mirrors that found in the 143 CyWT control and again strengthens the conclusion that there has been little impact on complex I assembly, despite reduced flavin availability.

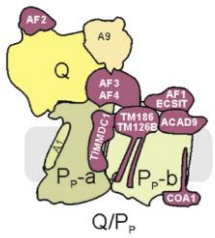
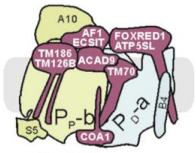
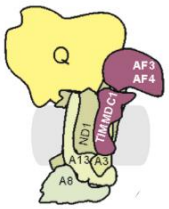
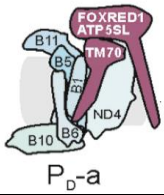
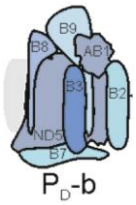
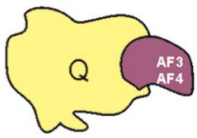
143B CyWT					
	Structure	**Estimated mass (kDa)**	**Calculated mass (kDa)**	**Subunits**	**Assembly Factors**
Subcomplex 1	 Q/P_P	650-719	680	NDUFA9, A5, S2, S3, S7, S8, ND1, A3, A8, A13, ND6, *ND4L*, ND3, ND2, *NDUFC1*, C2	NDUFAF2, NDUFAF3, NDUFAF4, TIMMDC1, TMEM126B, TMEM186, COA1, ACAD9, ECSIT, NDUFAF1
	 P_P-b/P_D-a		736	ND6, *ND4L*, ND3, ND2, *NDUFC1*, C2, A10, S5, B6, B5, B10, B11, B1, *ND4*, B4	TMEM126B, *TMEM186*, COA1, ACAD9, ECSIT, NDUFAF1, TMEM70, FOXRED1, *ATP5SL*
Subcomplex 2	 Q/P_P-a	292-357	283	NDUFA5, S2, S3, S7, S8, ND1, A3, A8, A13	NDUFAF3, NDUFAF4, TIMMDC1
	 P_D-a		230	NDUFB6, B5, B10, B11, B1, *ND4*	TMEM70, FOXRED1, *ATP5SL*
	 P_D-b		154	*NDUFAB1*, B7, B3, B8, ND5, B9, *B2*	-
Subcomplex 3		161	170	NDUFA5, S2, S3, S7, S8	NDUFAF3, NDUFAF4

Table 3.1 Possible subcomplexes present in 143B CyWT membrane complexomes. Estimated mass and subunit composition were used to assign possible identities to each subcomplex, using the most recent detailed complex I assembly paper (Guerrero-Castillo *et al*., 2017). Table shows all subunits and assembly factors expected to be present in each subcomplex, those in blue and underlined were present in the slice, black denotes detected subunits not present in the slice and those in italics were not detected in the complexome.

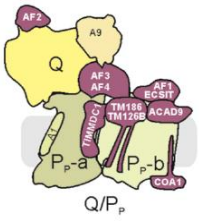
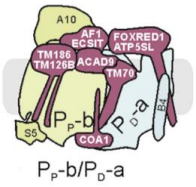
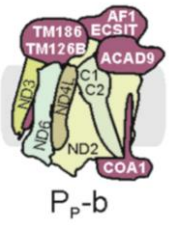
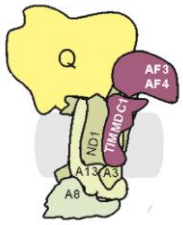
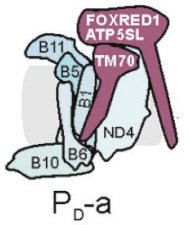
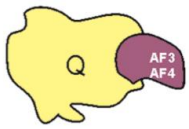
RFK KD					
	Structure	**Estimated mass (kDa)**	**Calculated mass (kDa)**	**Subunits**	**Assembly Factors**
Subcomplex I		732-799	680	NDUFA9, A5, S2, S3, S7, S8, ND1, A3, A8, A13, ND6, *ND4L*, ND3, ND2, *NDUFC1*, C2	NDUFAF2, NDUFAF3, NDUFAF4, TIMMDC1, TMEM126B, *TMEM186*, *COA1*, ACAD9, ECSIT, NDUFAF1
			736	ND6, *ND4L*, *ND3*, ND2, *NDUFC1*, C2, A10, S5, B6, B5, B10, B11, B1, ND4, B4	TMEM126B, TMEM186, *COA1*, ACAD9, ECSIT, NDUFAF1, TMEM70, FOXRED1, *ATP5SL*
Subcomplex 2		302-330	292	ND6, *ND4L*, *ND3*, ND2, *NDUFC1*, C2	TMEM126B, *TMEM186*, *COA1*, ACAD9, ECSIT, NDUFAF1
			283	NDUFA5, S2, S3, S7, S8, ND1, A3, A8, A13	NDUFAF3, NDUFAF4, TIMMDC1
			230	NDUFB6, B5, B10, B11, B1, ND4	TMEM70, FOXRED1, *ATP5SL*
Subcomplex 3		165	170	NDUFA5, S2, S3, S7, S8	NDUFAF3, NDUFAF4

Table 3.2 Possible subcomplexes present in the RFK KD membrane complexome. Estimated mass and subunit composition were used to assign possible identities to each subcomplex, using the most recent detailed complex I assembly paper (Guerrero-Castillo *et al*., 2017). Table shows all subunits and assembly factors expected to be present in each subcomplex, those in blue and underlined were present in the slice, black denotes detected subunits not present in the slice and those in italics were not detected in the complexome.

Whilst most of the subcomplexes were membrane spanning as expected in membrane fractions the presence of Q-domain intermediates was perhaps unexpected. The assembly pathway described by Guerrero-Castillo *et al.* was investigated using solubilised whole mitochondria and so this was not a factor in their data. In the complexome data above this could explained by matrix contamination or some association of the Q-domain intermediates with the membrane, possibly by the action of the assembly factors, to bring this domain nearer to its final point of assembly, on the forming complex in the inner mitochondrial membrane.

3.9. Mitochondrial matrix complexome analyses of RFK knockdown cells

The mitochondrial matrix complexomes were assessed for complex I subunits – whether subunits were present as free monomers or small soluble assembly intermediates (data for proteins given in Appendix 7.2F&I). NDUFV1 is an N-domain component, which sits in the matrix, and so changes to flavin incorporation may affect N-domain assembly. In the 143B CyWT matrix sample four subunits were detected, NDUFV1, NDUFV2, NDUFS1 and NDUFA2 – all part of the N-domain – along with the assembly factors NDUFAF5 and NDUFAF2. In the RFK KD matrix complexome these and a further six subunits were present (NDUFA5, NDUFA13, NDUFS2, NDUFS3, NDUFS8 and NDUFAB1). All except NDUFA13 (which spans the membrane and extends into the hydrophilic domain) and NDUFAB1 (with two copies in different locations) are found in the hydrophilic Q-domain of complex I.

NDUFA5, NDUFS2, NDUFS3, and NDUFS8 should assemble together as the core Q-domain; yet from Figure 3.27 it is clear that the profiles of these subunits demonstrated little cooperation in the RFK knockdown. Some subunits, particularly NDUFS2 and NDUFS3 had peaks in intensity at molecular weights suggesting multi-protein assemblies, whilst others suggested monomeric protein. However in both complexomes the profiles of NDUFV1 and NDUFV2 clustered together at lower molecular weights, which may indicate an association between the two subunits –supported by the assembly pathway from Guerrero-Castillo *et al.* However RFK knockdown and 143B CyWT matrix complexome samples were found to differ considerably in protein content – this indicated high levels of contamination, which is more likely in matrix compared to membrane fractions, and carry over from mitochondrial membranes and other cellular components during preparation. Therefore the samples were not consistent or homogenous and consequently not compared further. The calibration curve and table of proteins used to estimate the mass of proteins in the RFK knockdown matrix sample is given in Appendix 7.2G&H.

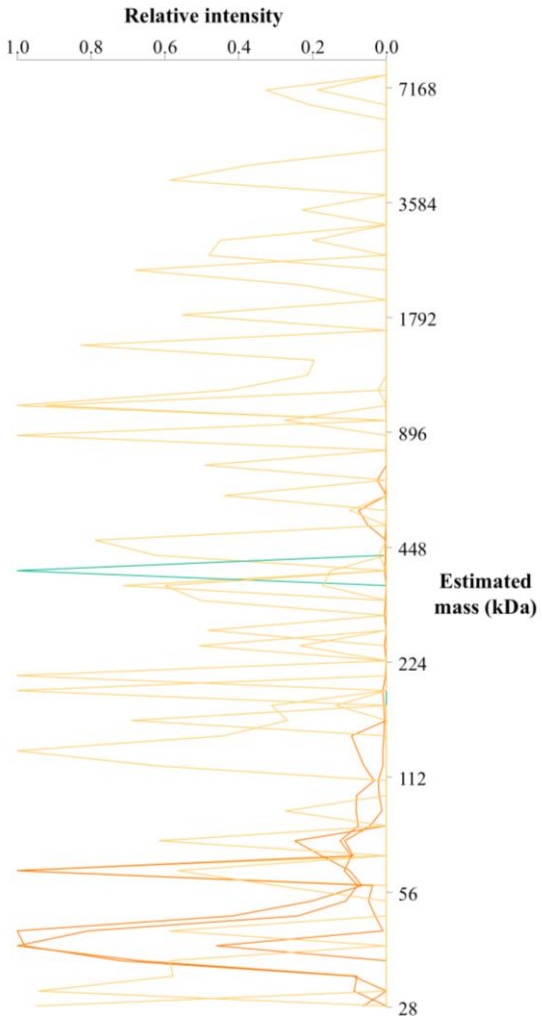


Figure 3.27 Clustering of complex I subunits in the RFK KD matrix complexome do not indicate the presence of soluble subassemblies. The 10 complex I subunits detected in the RFK KD matrix fraction showed differing profiles. Subunits are coloured by structural domain, yellow for Q-domain, orange for N-domain and green for hydrophobic domain.

NDUFV1 and NDUFV2 associate from the early stages of assembly; therefore other proteins with a similar migration profile might also interact with NDUFV1 and NDUFV2. 19 proteins were found and investigated for known function and candidacy as flavin binding/inserting proteins. Of these proteins 14 were non-mitochondrial and discounted, leaving five possible interactors: PDHX, GRSF1, REXO2, OXSM and SDHAF2. GRSF1 and REXO2 are nucleotide-binding enzymes with well-defined roles in mitochondrial gene expression. PDHX anchors components of the pyruvate dehydrogenase complex and OXSM is an acyltranferase, possibly involved in lipoic acid biosynthesis. None of these protein have flavin-binding motifs or chaperone characteristics and so were unlikely to be involved in the flavination of NDUFV1. However the complex II assembly factor SDHAF2, also known as PGL2 or SDH5, has been characterised as the FAD insertion factor for SDHA and the mutation G78R abolished interaction with, and flavin insertion into, SDHA (Hao *et al*., 2009). The requirement for FAD insertion into SDHA by a chaperone would indicate that a similar assembly factor may be required by NDUFV1. Additionally a shared assembly factor for two flavin-containing respiratory complexes could provide a sensing or regulating point through a single protein – due to the necessity for the flavin factor as the point of electron entry in each enzyme.

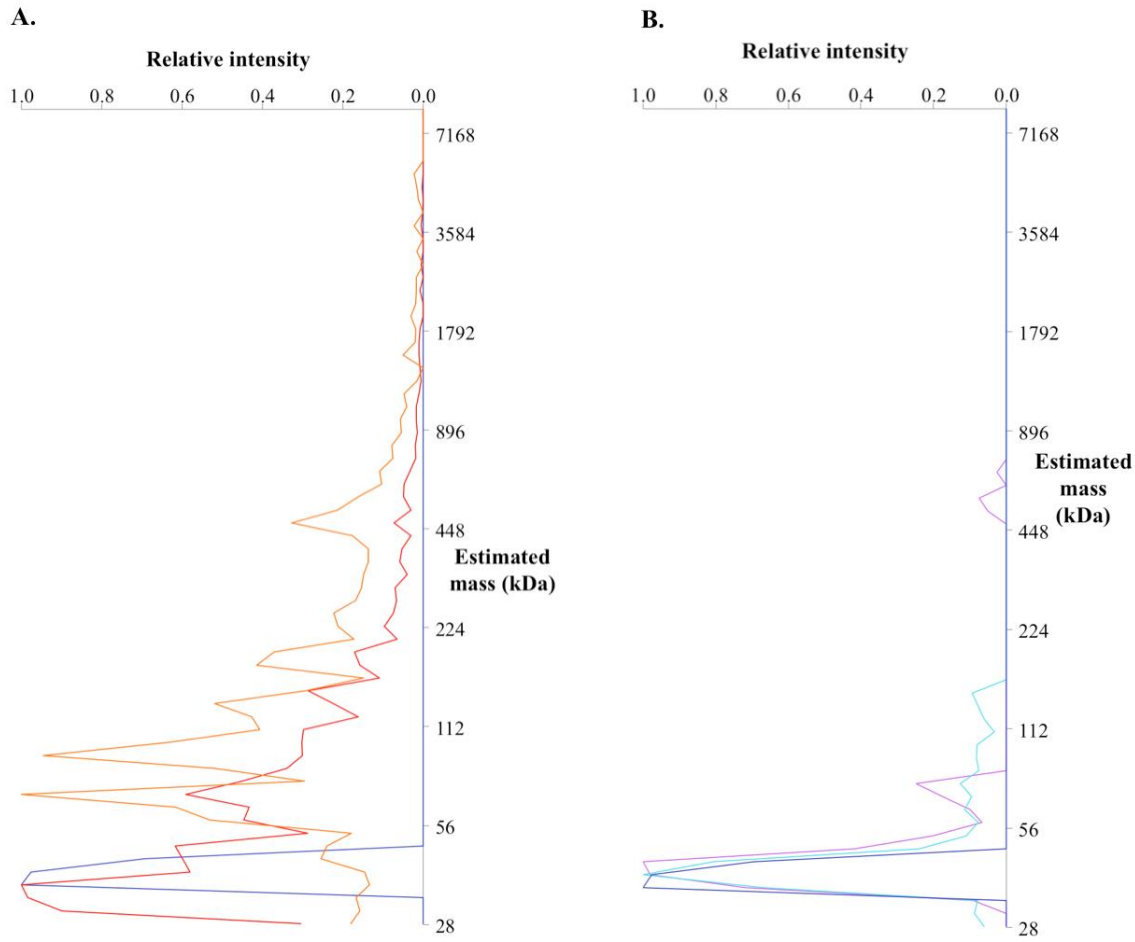


Figure 3.28 Matrix complexome profiles from RFK KD showing possible association between SDHAF2 and flavin containing subunits of complex I and complex II. Mitochondria were purified and matrix and membranes separated by sonication and after solubilisation proteins were separated by BN-PAGE, to maintain native complexes. **A.** SDHAF2 clustering with SDHA and SDHB from complex II. SDHAF2 shown in dark blue, SDHA in red and SDHB in orange. **B.** SDHAF2 clustering with NDUFV1 and NDUFV2 from complex I. SDHAF2 shown in dark blue, NDUFV1 in pink and NDUFV2 in light blue.

Figure 3.28A and B show the profiles for SDHAF2 overlaid onto the complex I subunits NDUFV1 and NDUFV2 or the complex II subunits SDHA and SDHB in the RFK KD matrix sample. The highest relative intensity of SDHAF2 corresponded with SDHA, as expected – but it also matches the single shared peak of NDUFV1 and NDUFV2 migration. Although an interesting observation, the low resolution of proteins at low molecular weight is a limitation of BN-PAGE. Therefore, from complexomics alone, an interaction between SDHAF2 and NDUFV1/NDUFV2 could not be validated.

Overall, complexomics provided a useful tool for assessing the structure and assembly of complex I in the 143B RFK knockdown. Comparison to the 143B CyWT demonstrated there were no striking differences in either complex I or other components of the OXPHOS chain. Lower molecular weight subcomplexes were detected as expected in the membrane fraction and the presence of assembly factors and supercomplexes were comparable to the control. Although the matrix fraction may have provided more detail on changes to the N-domain, the results demonstrated the high variation between matrix

preparations – exemplified by the migration of ribosomal proteins – which were found at smaller molecular weights in the knockdown but as multimers in the 143B CyWT, making comparison with the control unsuitable. Therefore while complexomics is a valuable technique, fractionation into matrix and membrane fractions may be unnecessary, as it does not necessarily enhance the data gathered.

3.10. SILAC experiments demonstrated changes to the mitochondrial proteome in response to flavin depletion

RFK knockdown coupled with flavin restriction was expected to severely affect the biosynthesis of flavin-containing proteins. At first it was not possible to source riboflavin-free SILAC DMEM – to allow heavy labelled amino acids (^{15}N) to be added and so a riboflavin rich control was used – 143B CyWT grown in 1 μM riboflavin. For this reason the initial data set used conditioned RFK KD 143Bs (un-labelled) and 143B CyWT grown in standard DMEM (heavy-labelled). Later, once riboflavin-free SILAC DMEM had been custom synthesised (Chapter 2, section 2.7.1), the data set was completed by reversing the labelling orientation.

RFK KD and 143B CyWT cells were processed simultaneously by mixing harvested cells at 1:1 protein ratio and then mitoplasts were prepared - along with a heavy control sample to check ^{15}N Arg and Lys incorporation. Samples were then reduced and alkylated before separation by SDS-PAGE and subsequent analysis by mass spectrometry (Chapter 2, section 2.7). The data discussed below are the total output from Perseus after processing of MaxQuant SILAC peptide pair analysis of RFK KD and 143B CyWT mitoplasts. Forward refers to heavy-labelled control mixed with un-labelled RFK KD; the reverse is the opposite labelling orientation (RFK KD heavy-labelled, 143B CyWT un-labelled). The Perseus processing stages included removal of proteins with less than one unique peptide or with a peptide count less than two; the normalised protein SILAC ratios were then presented as base two logarithmic. Forward orientation SILAC was repeated thrice and in the reverse orientation twice to provide a sufficient data set to obtain averages for each orientation and demonstrate the changes in protein abundance between conditioned RFK KD cells and the 143B CyWT control. Figures below display the Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation represent proteins with higher abundance in the 143B CyWT control (in line with the forward ratio) and negative values indicate increased abundance in the RFK KD. Those proteins with a three-fold or greater change in abundance in both labelling orientations were deemed considerably different between the samples (Log_2 ratio ≥ 1.58) and those with a greater than two-fold change highly changed (Log_2 ratio ≥ 1.00).

Overall 2980 proteins were detected, with 25 showing a greater than three-fold decrease in the conditioned RFK KD cells compared to the 143B CyWT control (Appendix 7.3). Of these a high

proportion were found to be particularly interesting, seven were subunits of the respiratory complexes and four were mitochondrial flavoproteins.

3.10.1. SILAC revealed changes in the abundance of complex I subunits located in the hydrophilic arm in RFK knockdown cells

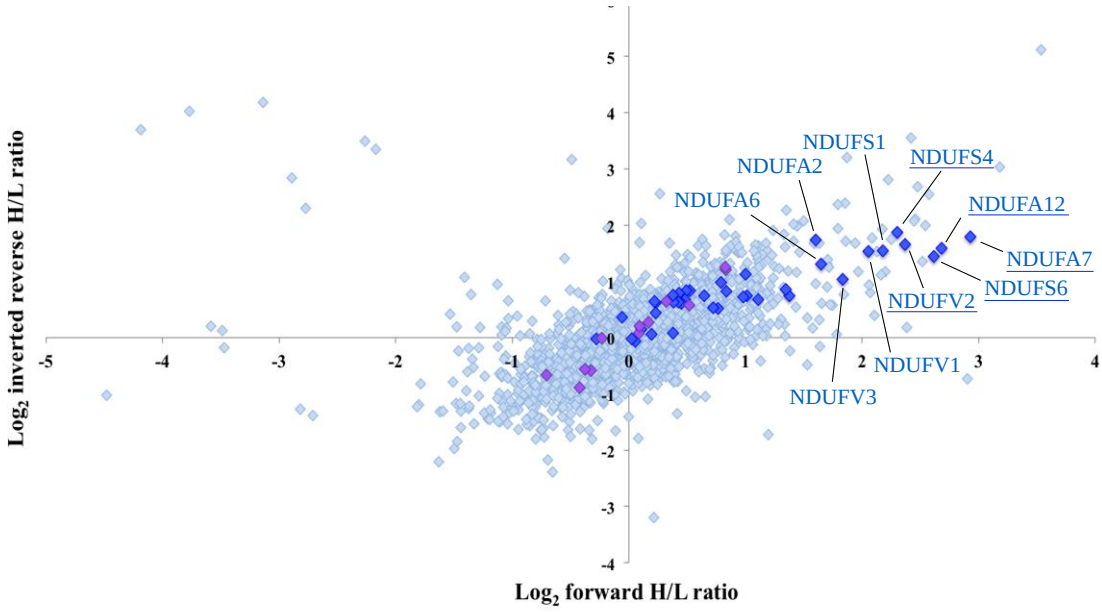


Figure 3.29 SILAC data comparing protein abundance from conditioned RFK KD 143Bs and 143B CyWT. Data is displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the 143B CyWT control and negative values indicate increased abundance in the RFK KD. Complex I subunits are highlighted in dark blue and assembly factors in purple. Complex I subunits with greater than a two-fold decrease (in both labelling orientations) are labelled and those with more than a three-fold decrease are underlined. Average data from three forward orientation (143B CyWT heavy labelled as "forward" orientation) and two reverse labelled experiments.

The SILAC data demonstrated good coverage of complex I subunits, with 40/45 detected (NDUFAB1 present as two copies, and ND2, ND4L, ND6, NDUFA1 and NDUFC1 were absent, Appendix 7.3). 29 subunits showed little difference in abundance, with less than a two-fold difference, although most H/L ratios were positive, indicating some decrease in abundance in the RFK KD cells. However, as shown in Figure 3.29, ten subunits showed a distinct difference between the cell types, five with a greater than two-fold decrease in the RFK knockdown - NDUFA2, NDUFA6, NDUFV3, NDUFV1 and NDUFS1. Additionally NDUFS4, NDUFV2, NDUFS6, NDUFA12 and NDUFA7 showed a more pronounced decrease, greater than three-fold, suggesting that whilst not all complex I subunits were affected a specific subset displayed markedly lower levels in the conditioned RFK KD. Figure 3.30 shows that those with a greater than three-fold decrease (green) and greater than two-fold decrease (magenta) are located at the top of the hydrophilic domain – the N-domain. Eight subunits comprise this region, responsible for NADH oxidation using the FMN cofactor, and all were found within the ten most changed subunits. NDUFA6 and NDUFA7 were also less abundant; these subunits form the top of the Q-domain, sitting adjacent to the N-domain of complex I. Therefore the data suggest that the FMN-

containing region of complex I is negatively impacted by RFK knockdown under riboflavin depleted growth conditions, such that there was a greater than two-fold decrease in these subunits, whereas the remainder of the enzyme was largely unaffected. Notably the abundance of Q-domain subunits NDUFA5, NDUFS2, NDUFS3, NDUFS7 and NDUFS8 are not markedly different, only the distal region of the hydrophilic arm was affected.

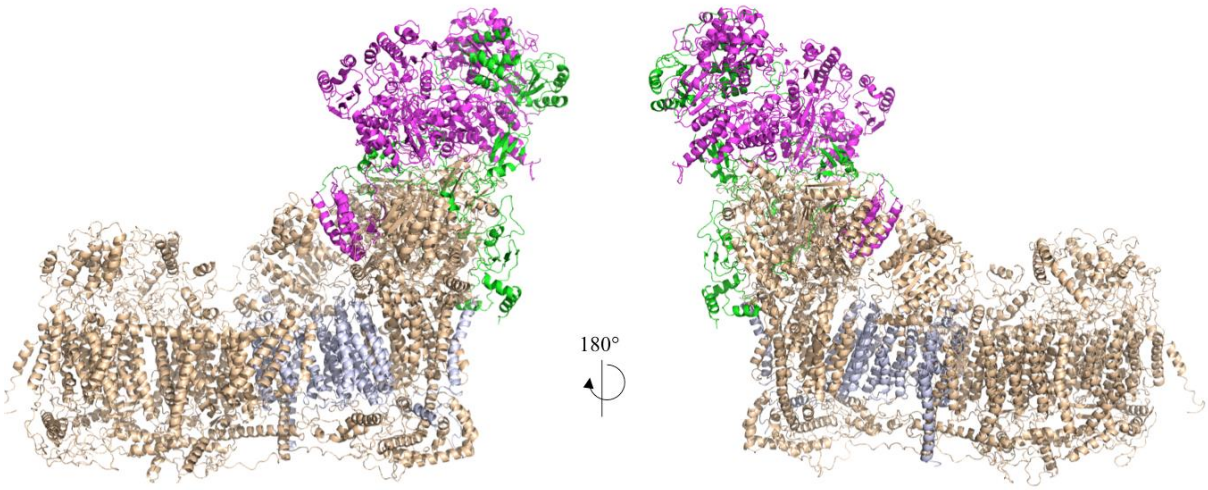


Figure 3.30 Complex I subunits showing the lowest abundance in RFK knockdown 143Bs highlighted on the humanised mouse model structure. Those demonstrating a greater than two-fold decrease are shown in magenta (NDUFA2, NDUFA6, NDUFV3, NDUFV1 and NDUFS1) and greater than a three-fold decrease in green (NDUFS4, NDUFV2, NDUFS6, NDUFA12 and NDUFA7). Other subunits are shown in wheat, with those not detected shown in grey. The human model of complex I was constructed using mouse complex I cryo-EM structures (Chapter 6, section 6.6).

Interestingly eleven assembly factors were detected, all of which showed little change (less than two-fold) in abundance. FOXRED1 and ATP5SL were the most different between the cell lines, and along with TIMMDC1, TMEM126B, NDUFAF1, ACAD9 and ECSIT, showed deceased abundance in the RFK KD, although not considerably. In contrast NDUFAF2, NDUFAF3, NDUFAF4 and COA1 displayed some increase in abundance, with COA1 showing >1.5 fold increase in the conditioned RFK KD. However from Figure 3.29 it can be seen that these differences are comparable to the majority of proteins detected – centring around $Log_2 = 0$ - and so may not be significant.

3.10.2 SILAC shows decreased abundance of complex II subunits in RFK knockdown cells

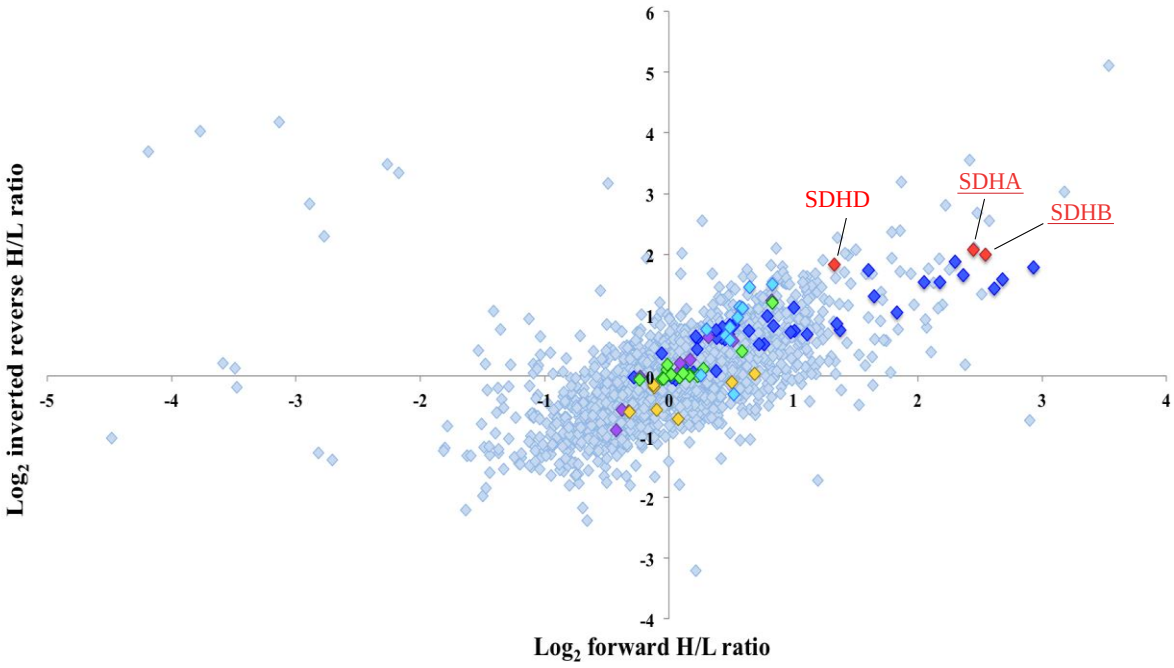


Figure 3.31 SILAC data comparing protein abundance from conditioned RFK KD 143Bs and 143B CyWT. Average data from three forward and two reverse labelling orientation repeats. Data is displayed as Log_2 H//L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the 143B CyWT control and negative values indicate increased abundance in the RFK KD. Complex I subunits are highlighted in dark blue and assembly factors in purple, complex II in red, complex III in yellow, complex IV in green and complex V in light blue. Complex II subunits with greater than a two fold decrease (in both labelling orientations) are labelled and those with a greater than three-fold decease are underlined.

The abundance of subunits from complexes II-V were also analysed to assess whether RFK KD and riboflavin depletion affected the overall levels of mitochondrial respiratory complexes (Appendix 7.3). Figure 3.31 shows eight complex III subunits (yellow), 11 complex IV subunits (green) and 16 complex V subunits (light blue) were detected – all demonstrating a less than two-fold change in abundance in both orientations. In contrast the complex II subunit SDHD was reduced by more than 2.5 fold, with SDHB and the FAD containing subunit SDHA decreased by more than four-fold in conditioned RFK knockdown cells (red). This suggests that, whilst there was no universal decrease in the oxidative phosphorylation complexes, the abundance of complex II was greatly decreased, similar to the lowered abundance of the N-domain subunits of complex I. It could be speculated that as complex II is a smaller enzyme (130 kDa) than complex I (~1 MDa) the affect of low flavin would impact the stability of the entire protein, such that all four subunit abundances would be decreased (SDHC was not detected). In contrast complex I, with 45 subunits and the FMN domain at one end of the structure, may not be globally affected.

3.10.3. SILAC highlights a hierarchal change in mitoflavoprotein abundance in RFK knockdown cells

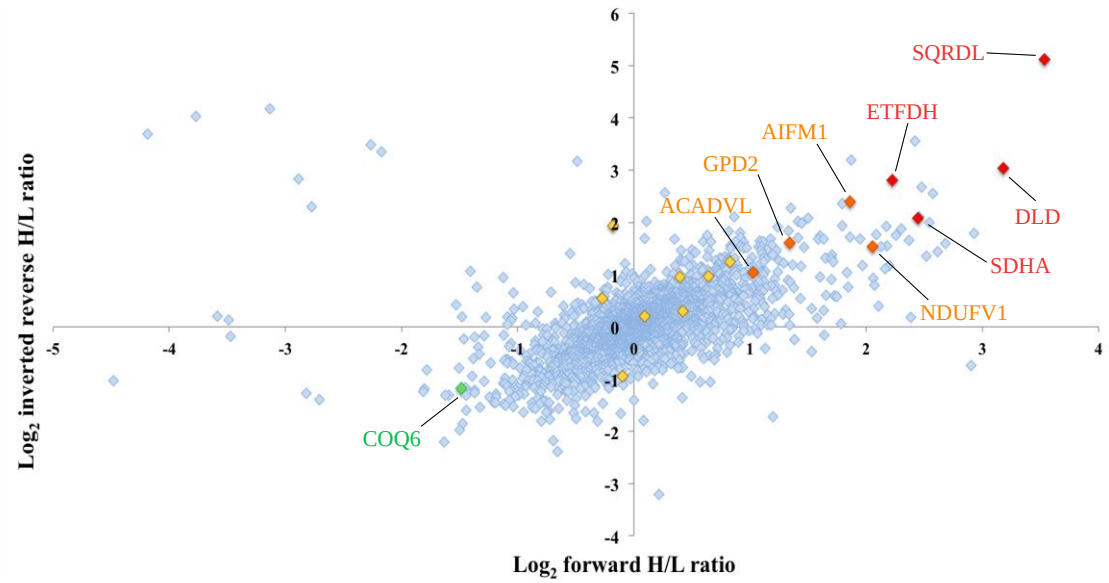


Figure 3.32 SILAC data comparing protein abundance from conditioned RFK KD 143Bs and 143B CyWT. Average data from three forward and two reverse labelling orientation repeats. Data is displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the 143B CyWT control and negative values indicate increased abundance in the RFK KD. All flavoproteins detected are coloured yellow, those with a greater than two-fold decrease (in both labelling orientations) are shown and labelled in orange and a greater than three-fold decrease in shown and labelled in red. Those increased by more than two-fold are highlighted and labelled in green.

There are 32 flavoproteins located within mammalian mitochondria requiring either FMN or FAD as redox cofactors (Lienhart *et al.*, 2013). 18 were detected in the SILAC data (Appendix 7.3) and analysis suggested that mitoflavoproteins reacted hieratically to flavin starvation. Although it may be hypothesised that all flavin-containing proteins would be detrimentally affected this was not reflected by the data (Figure 3.32). Instead 50% were considered unchanged and 28% demonstrated a considerable decrease of more than three-fold in the RFK KD.

Four flavoproteins were highly decreased in conditioned RFK KD 143Bs: SQRDL (> 11-fold), DLD (>8-fold), SDHA (>4-fold) and electron-transfer dehydrogenase (ETFDH, >7 fold). To gain insight into this apparent flavoprotein hierarchy, with some proteins more detrimentally affected than others, the function of these proteins and possible links between them were investigated using the literature.

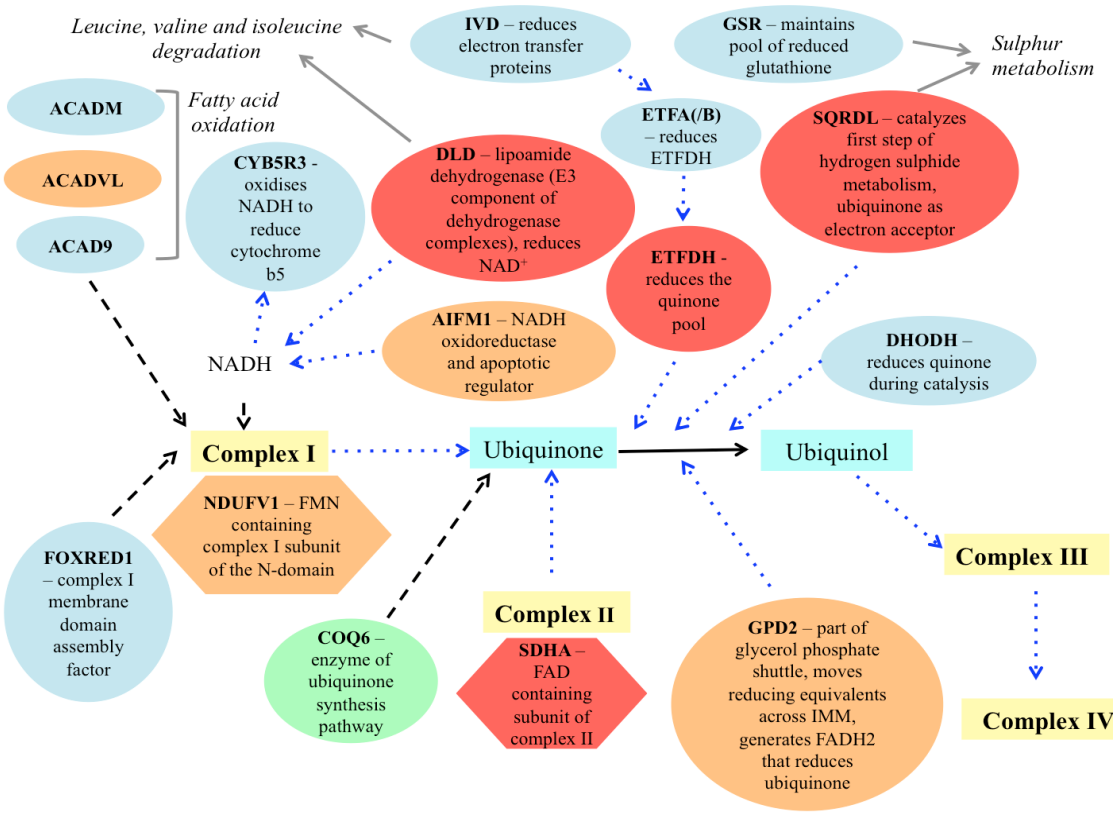


Figure 3.33 Diagram exploring the functions and pathway connections between flavoproteins detected in SILAC data. Flavoproteins are coloured by the extent of change demonstrated by SILAC – for the RFK KD compared to the control: red for greater than three-fold decrease, orange for greater than two-fold, blue for unchanged and green for greater than a two-fold increase in abundance. Blue dashed arrows reflect electron transfer and black dashed arrows highlight linked proteins (such as complex I assembly factors). Hexagons show the flavoprotein subunits of complex I and II. Other pathways are indicated in grey. The schematic was arranged around ubiquinone reduction as many of flavoproteins are linked to this pathway either by directly transferring electrons or interaction with the respiratory chain.

Figure 3.33 classifies the 18 mitoflavoproteins by function and highlights that the beginning of the electron transfer chain and reduction of ubiquinol are key. Flavoproteins are coloured by the extent of change demonstrated by SILAC. Although not all flavoproteins were directly linked to OXPHOS (more distant pathways are indicated in grey), there was a striking clustering of the most decreased proteins, due to their roles in the reduction of ubiquinone and NAD^+. Additionally NDUFV1 and GPD2 also have close links with this pathway and were more than two-fold decreased in the RFK KD. However a detrimental effect on proteins with a central role in maintaining the redox status of NADH and ubiquinone was unexpected – as a flavoprotein hierarchy would be expected to prioritise important proteins to receive their cofactor, leaving more redundant flavoproteins to cope with the biggest flavin reduction. Instead, as highlighted visually in Figure 3.33, most proteins demonstrating a large decrease in the conditioned RFK KD had a direct link to the ubiquinone pool or NADH/NAD^+ ratio, with those least affected involved in other pathways in the mitochondria.

This may be explained in part by the basal abundance of each flavoprotein, with some in large excess though still capable of adequate function at lower abundance, whereas others may be intolerant of small

changes. The apparent flavin hierarchy may also reflect the stability of these proteins and the importance of the flavin cofactor in their structural integrity. However it cannot be determined whether these differences in abundance are a direct or indirect result of decreased flavin – lowered abundance may be due to the inherent instability of flavoprotein lacking its cofactor, or may be a coordinated feedback response from the cell to lower expression of certain flavoproteins, in response to limited cofactor availability. To delineate this a complementary approach looking at expression levels would be required, such as RNA transcriptome profiling.

A further curiosity was the considerable increase of one mitoflavoprotein; COQ6 is a component of the multi-subunit COQ complex, with a role in ubiquinone synthesis. It has an FAD cofactor and interacts with five other COQ proteins (COQ3-7 and 9) and was increased by more than two-fold in conditioned RFK KD cells, along with COQ3, 5, 7 and 9 to a lesser extent. This would suggest that, despite being a flavoprotein, levels of COQ6 and other members of the ubiquinone synthesis pathway, were increased in response to changes in the redox state of the ubiquinone pool – possibly in an attempt to increase the amount of ubiquinone and offset changes to its redox state. Consequently the SILAC data obtained demonstrated changes in protein abundance directly due to low flavin, in addition to those instigated by the cell in response to this stress.

3.11. An intermediate condition determined that riboflavin depletion in the growth media was the dominant cause of the changes in protein abundance

The initial SILAC analysis used a control cell line grown in standard, riboflavin-rich DMEM, due to the unavailability of riboflavin-free SILAC DMEM. However it later became possible to obtain the riboflavin-free SILAC DMEM, thus to delineate the effects of RFK knockdown from flavin depletion an intermediate sample was prepared – conditioned 143B CyWT cells expressing the empty pLKO.1 vector. Two more pairs of cell lines were therefore processed for SILAC analysis:

> Pair 1: 143B CyWT in standard DMEM and conditioned pLKO.1 143B CyWT cells
> *(two repeats in forward labelling orientation, with 143B CyWT heavy labelled – "control", and two in the reverse labelling orientation).*
>
> Pair 2: conditioned pLKO.1 143B CyWT cells and conditioned RFK knockdown 143B cells
> *(two repeats in forward labelling orientation, with conditioned pLKO.1 143B CyWT heavy labelled – "control", and one in the reverse labelling orientation)*

Pair 1 represented the impact of riboflavin depletion alone, whereas pair 2 showed the additional influence of the RFK knockdown on cells already coping with restricted flavin. The data obtained were

from a smaller set of repeats than SILAC from section 3.10 and the quality was therefore lower, with decreased protein coverage.

Pair 1 - 143B CyWT/conditioned pLKO.1 cells - showed that 5.6% of proteins detected were decreased in abundance by more than two-fold in the conditioned pLKO.1 143Bs, demonstrated by the spread of data in Figure 3.34 (Appendix 7.4). Additionally Figure 3.34B shows that many mitoflavoproteins were decreased by more than three-fold under riboflavin starvation – including NDUFV1, SDHA, SQRDL, DLD and ETFDH, which were all considerably decreased in the 143B CyWT/conditioned RFK KD pair discussed previously (section 3.3.4.1). All proteins that showed a greater than two-fold decrease in section 3.10 also showed a comparable effect in this experiment – in most cases the decrease was more pronounced. However COQ6 did not demonstrate the same increase in abundance. The complex II subunit SDHA and ten complex I subunits were more than three-fold decreased, reflecting the large changes observed with the 143B CyWT/conditioned RFK KD pair. This would suggest that the results obtained from conditioned RFK KD 143Bs, with a control grown in riboflavin-rich conditions, were mirrored by riboflavin depleted growth conditions (pair 1), and therefore the effect was independent of the RFK knockdown. Additionally the differences in protein abundance shown by the conditioned cells compared to those grown in riboflavin-rich conditions (pair 1) is similar to the SILAC data discussed in section 3.10, which showed that a shorter period of more severe riboflavin depletion also caused changes in complex I subunit and mitoflavoprotein abundance. Therefore whilst conditioned 143B CyWT cells remain viable and had no clear phenotype differences in the mitochondrial proteome did develop, compared to when riboflavin was abundant.

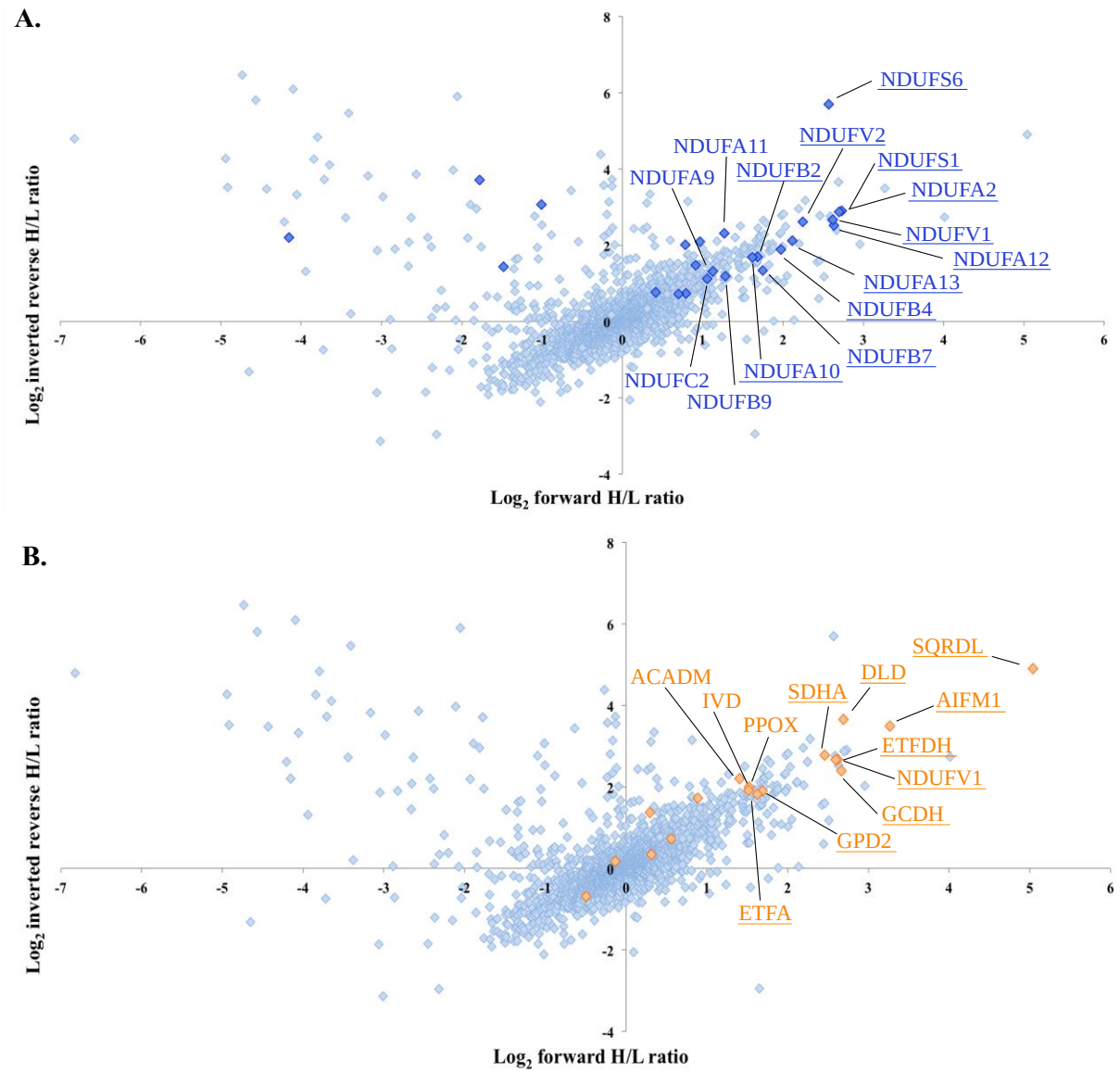


Figure 3.34 SILAC data for pair 1 to compare protein abundance from 143B CyWT cells grown in riboflavin rich and riboflavin depleted conditions. Average data from two forward and two reverse labelling orientation repeats. Data is displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the 143B CyWT control. **A.** Complex I subunits are highlighted in dark blue and those with above a two-fold decrease in abundance (in both labelling orientations) after conditioning are labelled (underlined showed a greater than three-fold decrease). **B.** Mitoflavoproteins are highlighted in orange and those with above a two-fold decrease in abundance (in both labelling orientations) after conditioning are labelled (underlined showed a greater than three-fold decrease).

For completeness pair 2 - conditioned pLKO.1 143B/conditioned RFK KD 143B CyWT cells - was analysed to compare the effect of RFK knockdown on cells already coping with riboflavin depletion.

Overall there was less change in protein abundance, with no proteins showing a greater than two-fold decrease in the RFK KD. Figure 3.35 shows that the abundance of most proteins were unchanged between the samples, including the 12 complex I subunits and 9 mitoflavoproteins detected (Appendix 7.5). Although the data was of lower quality than pair 1 (1218 proteins detected in total compared to 2090), the overall spread of data suggested little change and provided a good indication that RFK knockdown did not exacerbate the proteomic differences caused by conditioning the cells.

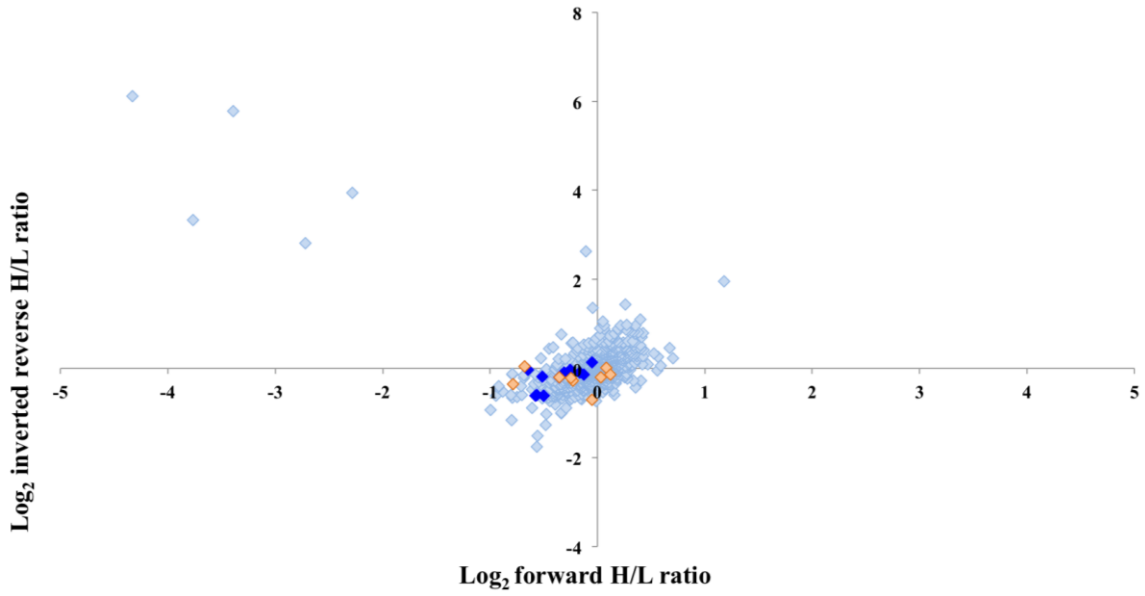


Figure 3.35 SILAC data for pair 2 to compare protein abundance of conditioned pLKO.1 to conditioned RFK KD 143B cells. Average data from two forward and one reverse labelling orientation repeats. Data is displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio, such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the conditioned pLKO.1 control. Complex I subunits are highlighted in dark blue and mitoflavoproteins in orange.

It was concluded that the results from SILAC, notably changes in certain flavoproteins – demonstrated a hierarchy, and decreased levels of complex I N-domain subunits were caused by the riboflavin depletion in the growth media and were not enhanced by the RFK knockdown. This was surprising given the crucial role of the enzyme in flavin metabolism at the early stages of the flavin cofactor pathway. However the knockdown was assessed to only be 88%, therefore a residual 10% could be maintaining turnover of the riboflavin required to synthesise flavin cofactors – and so there would be no difference from conditioned cells.

3.12. Discussion and conclusions

3.12.1. The importance of growth conditions – riboflavin excess

The work described in this chapter used the cultured human cell line 143B, widely used in mitochondrial studies. Manipulation of this cell line facilitated investigations into complex I assembly and flavin handing that would have been restricted by using other models, such as purified bovine material. Also, as a human system, the results obtained should more closely reflect processes occurring *in vivo*, in comparison to using genetically malleable microbial systems, such as *Y. lipolytica* and *P. denitrificans*. However, it must be considered that, whilst the biochemistry of cultured human cells resembles that within tissues, the characteristics making them suitable for prolonged culture also make them inherently different from those *in vivo*. 143Bs are derived from human osteosarcoma cells and so, as immortal cancer-like cells, they will have different metabolic adaptions; their response to stress and genetic manipulation may vary. Therefore, when interpreting results from using cultured cell lines this caveat must be considered. Whilst these cells are an incredibly useful tool for studying human defects, they cannot be directly paralleled with the processes of a whole organism.

The results presented in section 3.2 demonstrate that a dominant feature affecting flavin handling by 143Bs is the concentration of riboflavin in the growth media, which was found to be in large excess (standard DMEM has 1 μM riboflavin and 10% FBS contributed an additional 0.77 nM). Thus, even siRNA knockdowns of known riboflavin processing enzymes had a negligible effect on cellular flavin availability because high extracellular riboflavin concentrations compensated for decreased uptake and processing by the cells. Guided by previous work from Zempleni and co-workers, a conditioning period was adopted with growth of 143Bs in 0.77 nM riboflavin for one month before the cells were used. This consideration could be beneficial for other studies, to ensure cells are closer to the threshold for phenotypic changes. Riboflavin supplementation has been shown to improve the clinical symptoms of patients with respiratory defects and other conditions, such as multiple acyl-coA dehydrogenase deficiency, indicating that riboflavin can ameliorate these phenotypes (Greibel et al., 1990; Cotelli et al., 2012). Therefore, growth of patient fibroblasts in culture may not exhibit the same level of impairment if riboflavin is present at high levels and so this must be accounted for.

3.12.2. Complex I subunits and assembly in low flavin conditions

The aims of this chapter focused on the insertion of FMN into NDUFV1 and whether manipulating the upstream flavin-handling pathway could provide insight into the stage at which flavin is incorporated; including the importance of this for complex I assembly. Complexomics was used to investigate low flavin conditions and determine whether assembly factors were required for flavination of NDUFV1. Lowered riboflavin availability did not detrimentally affect complex I assembly – as shown by the

complexome analyses (section 3.3.3.1). Additionally, no candidate proteins were detected in association with NDUFV1, however the close association of NDUFV1 and NDUFV2 was noted because they shared the same migration profile in the mitochondrial matrix fraction, thus they must interact from a very early stage. The N-domain has been shown to arise separately during assembly (Guerrero-Castillo *et al*., 2017) but the association of these two subunits could also represent a stabilising strategy; NDUFV2 may support NDUFV1 before FMN is inserted. Assembly factors have been identified for flavoproteins requiring covalent attachment of their cofactor, such as SDHAF2 for FAD insertion into SDHA. However the interaction of non-covalent binding flavoproteins with flavin is usually high affinity in a shielded binding pocket (Starbird *et al*., 2015). Therefore, NDUFV1 may not require an insertion factor and instead obtain FMN readily from the matrix. Further work with a cell line demonstrating a severe respiratory phenotype would more conclusively establish this hypothesis, as conditioned RFK knockdown 143B cells still assembled active complex I.

Another consideration was highlighted by the work on the RFK knockdown. There was disparity between OCR measurements, complex I in-gel activity and complexomics compared to the SILAC proteomics data. There was no respiratory phenotype observed for either conditioned 143Bs or those with the RFK knockdown, suggesting that oxidative phosphorylation was not impaired by riboflavin depletion, and this was corroborated by the complexome data. Yet SILAC indicated a considerable decrease in the abundance of N-domain subunits from complex I and also a decrease in complex II subunits, for cells conditioning in low riboflavin. These data reflect the changes to mitochondrial proteins observed in section 3.4., for cells grown in 0.15 nM riboflavin that exhibited a severe resipiratory phenotype with reduced viability, and so it would be expected that a similar phenotype would be observed. The difference between SILAC and other experiments, in section 3.6 onwards, could indicate that a substantial pool of free complex I and II subunits exists and so despite a decrease in protein abundance under riboflavin restricted conditions (shown by SILAC) this was not to a level that impacted heavily on the assembly or the activity of the enzymes themselves, thus a phenotype did not manifest. However another plausible explanation involves differences in the growth conditions of these cells. Despite the same conditioning in all experiments, and subsequent processing to generate mitoplasts for analysis (for in-gel activity and SILAC), one element differed. Dialysed FBS was required for SILAC to prevent the introduction of unlabelled Arg and Lys into the heavy labelled cell sample, whereas usually all cell lines were cultured with untreated FBS. Although 10% FBS addition contains a low concentration of riboflavin (~0.77 nM) this will vary by batch, but importantly dialysed FBS contains much lower levels. Consequently the concentration of riboflavin available may be substantially decreased and could explain why the riboflavin depleted SILAC cells (with or without RFK knockdown, section 3.10) displayed highly changed levels of flavoproteins, because the concentration of riboflavin provided by 10% dFBS was lower than 10% untreated FBS, which pushed the cells beyond the threshold of tolerance, hence proteomic changes were evident. Thus under SILAC conditions these cells were more equivalent to the 0.15 nM riboflavin grown cells that were assayed previously, whereas in other

experiments the concentration of riboflavin was adequate for survival and functional OXPHOS. The impact of different FBS types has been found across different research areas, including receptor-ligand binding investigations for the pharmacological classification of ligands (Lau & Chang, 2014) and the sensitivity of cancer cells to bioenergetic inhibitors (Fan *et al*., 2010). Thus, when the precise growth conditions of cells must be controlled, the variation in small molecule concentrations in FBS can have an unintended but substantial effect.

3.12.3. Flavin depletion revealed a mitoflavoprotein hierarchy

From the SILAC data investigating the effects of flavin starvation on the mitochondrial proteome a differential change in mitoflavoprotein abundance was demonstrated. Not all flavin-containing proteins were affected to the same extent; some did not change in abundance whilst COQ6 was considerably increased. All of these proteins require a FAD cofactor, except NDUFV1 and dihydroorotate dehydrogenase (DHODH) that bind FMN. The hierarchy may result from a number of different factors. First it may indicate the strength of cofactor binding, the higher affinity proteins would monopolise the remaining flavin, although SILAC cannot distinguish between flavoproteins with or without their cofactor. Most flavoprotein cofactors are non-covalently bound but tightly associated; although it is estimated about 10% of flavoproteins contain covalently bound flavin and are more likely to rely on this association for structural stability (Heuts *et al*., 2009). Therefore a second factor influencing the mitoflavoprotein hierarchy is whether protein abundance is linked to protein stability – which in turn may be dictated by cofactor binding. It was shown that mutant monoamine oxidase (MAOA) that could not bind FAD covalently had lower stability upon purification (Nandigama & Edmonson, 2009). SDHA demonstrated a higher decrease in abundance compared to NDUFV1 in flavin-depleted cells, and this may be explained by their different flavin binding – SDHA covalently binds FAD, whereas NDUFV1 non-covalently associates with FMN. Therefore some flavoproteins, despite being unable to obtain their cofactor, persist for longer before being turned over compared to those that are inherently unstable without flavin. SILAC data may therefore highlight two properties of mitoflavoproteins that confer differing abundance during flavin starvation – protein-flavin binding affinity and structural integrity.

A final consideration influencing the mitoflavoprotein hierarchy is the involvement of additional proteins. The requirement for a flavin assembly factor adds an additional step in the flavination process, which in a flavin depleted environment may be detrimental. Also FAD must be synthesised from FMN, either in the cytoplasm or mitochondrial matrix by FADS isoforms, therefore availability of FAD will be limited by this enzymatic reaction – perhaps giving an advantage to FMN-containing flavoproteins. Although SILAC was only concerned with mitochondrial proteins it would be informative to compare the impact of flavin depletion on other flavoproteins, as flavin transport into organelles may affect cofactor availability between different compartments and subsequently affect flavoproteins. There is also the difference in gene or protein level effects, as SILAC provides no information on transcriptional

control, which would indicate a cellular feedback response to flavin starvation rather than just alterations to the proteins themselves.

3.12.4. Robustness of the flavin handling pathway and associated proteins

The surprising result was that knocking down RFK by ~90% did not further disrupt riboflavin handling, over conditioning alone, despite it being the only known riboflavin kinase. This suggested that the pathway is inherently robust and difficult to influence, which was also supported by the results presented in section 3.2, such that riboflavin concentrations in the growth media had to be lowered to ~0.15 nM to confer a phenotype on the cells, and the riboflavin uptake inhibitor lumiflavine was only effective in conjunction 0.77 nM riboflavin growth conditions.

3.12.5. Future developments

Overall, this investigation into flavin-handling by cultured cells and the impact on complex I activity and assembly produced a number of interesting results – including the lack of RFK knockdown phenotype, and that growth conditions must be addressed to decrease flavin flux within the cell. To take this project forward, proteins of interest would need to be knocked out. Thus the next stage would be to affect a genomic level knockout of RFK, along with the poorly characterised mitochondrial carrier SLC25A32, with a view to pushing the cell as a whole, and complex I in particular, towards a flavin starved state, to better study the process of flavin cofactor insertion. This should then allow a more detailed investigation into FMN incorporation into the crucial NDUFV1 subunit of complex I.

However, data from the Mouse Genome Informatics online resource showed mice homozygous for a RFK knockout allele died *in utero* (Yazdanpanah *et al*., 2009). This indicates the importance of the enzyme, although it must be noted that inviability of an embryo with a gene knockout only demonstrates a crucial developmental role for the protein. It may be possible to generate a knockout in cultured cells, when its developmental function is no longer required. This has been shown for many proteins, specifically of interest is the complex I subunit NDUFA5. A mouse knockout was generated using gene trap by Moraes and co-workers but was embryonic lethal and survival did not exceed day E9 (Peralta *et al*., 2014). However a conditional knockout in the central nervous system was viable, but a severe phenotype was observed in these mice at ~11 months. Additionally, Ryan and co-workers have knocked out all the supernumerary subunits of complex I and showed that in HEK293T lines only NDUFAB1 was essential for cell viability (Stroud *et al*., 2016). Therefore, although NDUFA5 knockout cells showed severely decreased complex I assembly, indicating the importance of the subunit in the biosynthesis and stability of the enzyme, these cells survived, unlike the knockout mice. Considering this the possibility of knocking out RFK in cell lines was explored. Horizon Discovery had previously used CRISPR to target the gene in HAP1 and other cell lines, but they were unable obtain viable clones

(*pers.comm.*). Therefore RFK may represent an essential protein in both development and cell survival, thus the absolute requirement for this enzyme may explain the lack of effect found on knockdown, ~90% decrease in protein abundance could represent the highest viable level of knockdown possible, and those with a higher knockdown efficiency cannot survive. Unfortunately this meant further work with RFK was not possible and so CRISPR knockout was pursued with SLC25A32 alone.

Chapter 4

Investigating the function of SLC25A32 and the effect of its knockout on complex I in cell lines

4.1. Introduction

4.1.1. Mitochondrial carriers

The two membranes of mitochondria exhibit differing permeability, with the OMM allowing movement of most small proteins and molecules due to the presence of porins and using translocases to import larger proteins. The IMM, in contrast, is tightly regulated and specific transporters such as the mitochondrial carrier family (MCF) are required for ion and metabolite movement. Figure 4.1 demonstrates the diverse substrates transported by mitochondrial carriers and the key biochemical pathways of the matrix – highlighting the role of MCFs in linking cellular and mitochondrial metabolism.

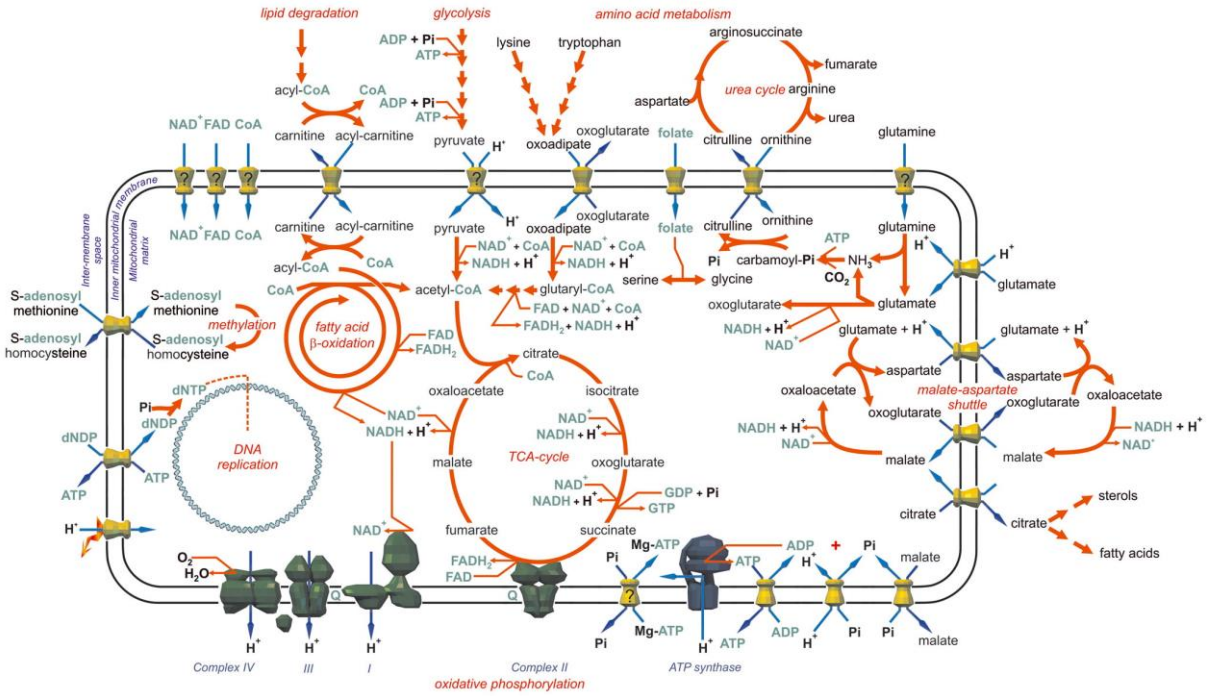


Figure 4.1 Mitochondrial carriers of the IMM. A schematic representation of the mammalian IMM respiratory chain complexes (grey) and mitochondrial carriers (yellow). The OMM has been omitted, orange arrows indicate simplified metabolic pathways, to emphasise key metabolites, with cofactors and high-energy intermediates in teal. Adapted from Kunji, 2004.

There are over 50 MCFs with a large range of substrates, although they all share defined features that are essential for their function. The sequences of MCF proteins show a tripartite structure of three

homologous sequence repeats comprising 100 residues. Each repeat consists of two transmembrane helices and a short helix that extends into the mitochondrial matrix. Within these repeats there is a conserved signature motif of P-X-[D/E]-X-X-[R/K] on one transmembrane helix. The carriers were originally thought to act as homodimers, with evidence supporting this including the 1:2 stoichiometry of the inhibitor carboxyatractyloside (CATR) binding to the ATP/ADP carrier (Riccio *et al*., 1975). However, this dimer theory was later dispelled by a series of investigations by Kunji and co-workers confirming that the ADP/ATP carrier is a monomer, using purification and functional assays with work on the yeast protein (Bamber *et al*., 2007).

The carriers adopt two states; one accessible to the matrix (m-state) and one open to the cytoplasm (c-state). The signature motif contributes charged residues to form a salt bridge network that closes the internal cavity to the matrix side. A further conserved motif was identified ([Y/F]-[D/E]-X-X-[K/R]) on the second transmembrane helix in each repeat that forms a salt bridge to close the cavity to the cytoplasmic side during the transport cycle (Crichton *et al*., 2015). Fixing the carrier in different states using inhibitors has provided valuable insights into its binding properties and allowed mechanisms to be proposed.

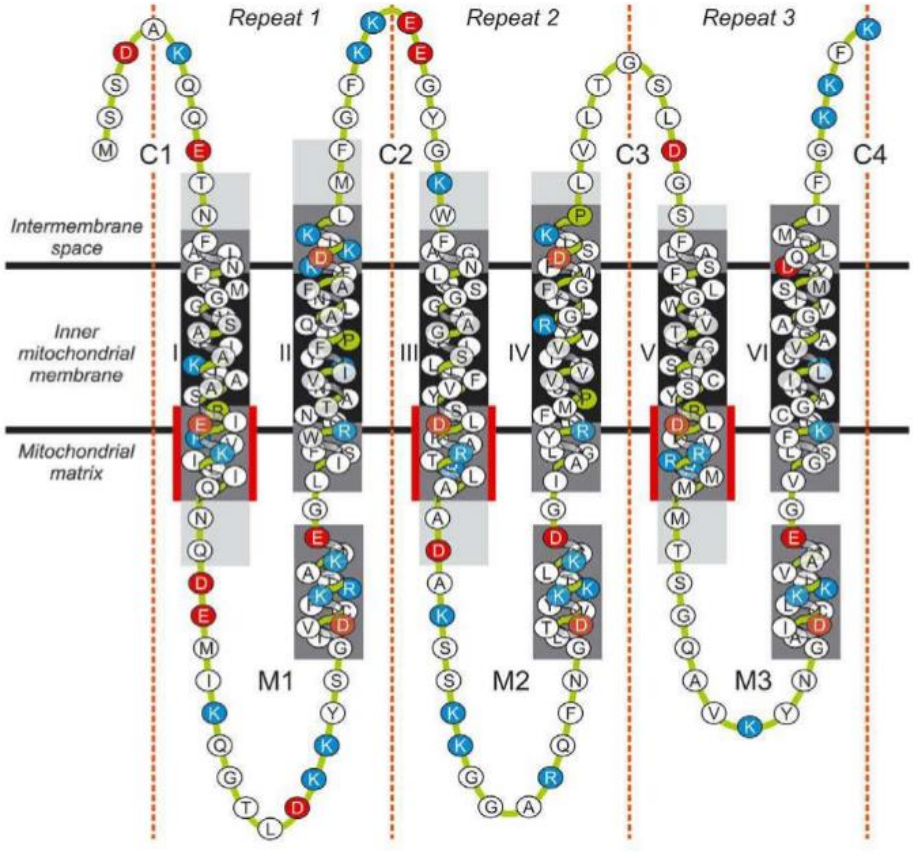


Figure 4.2 Topology of the yeast ADP/ATP carrier. The six transmembrane helices are shown in black, with red dashes separating the three sequence repeats. The position of the signature motif P-X-[D/E]-X-X-[R/K] of the first transmembrane helix is bordered in red. Negatively charge residues are shown in red and positively charged in blue. The matrix loops are numbered M1, M2 and M3, cytoplasmic loops C1, C2, C3 and C4. Adapted from Kunji, 2004.

Whilst identifying MCF members has been relatively straightforward due to their highly conserved motifs, determining their functions has proved challenging; to date only around a third have been assigned substrates by experimental evidence. The main approaches for characterisation involve expression in *E. coli* and *S. cerevisiae* and reconstitution into liposomes for direct measurements of substrate transport (Palmeiri and Pierri, 2009). More recently a *Lactococcus lactis* expression system has been used, allowing the carrier to be characterised in whole cells or membrane vesicles without additional purification and reconstitution steps (Kunji, 2004). Despite these approaches, the nature of these proteins as membrane-localised carriers has impeded successful expression and purification of many MCF members; they have been found to be unstable when purified in different detergents, hampering their functional assignment (Crichton *et al*., 2015). As a result many MCF proteins have been given a putative function based on computational prediction and observations in genetic models, including the rescue of mutations by expression of these carriers.

4.1.2. The putative mitochondrial folate carrier, SLC25A32

The function of SLC25A32 has divided opinion since two separate investigations into its physiological substrate have yielded different results. As discussed in Chapter 3, section 3.1.1. it was assigned as the folate transporter after rescue of glycine-auxotrophic Chinese hamster ovary-derived glyB cells by overexpression of the protein – the cause was later assigned to a mutation in the transporter (G192E) (Titus *et al*., 2000). However, in 2005 Spann and co-workers reported the rescue of a *S. cerevisiae* FLX1 null mutant (assigned as the mitochondrial FAD transporter) with expression of human SLC25A32, and so this indicated an ability to transport flavin (Spaan *et al.,* 2005). The debate has also been fuelled by the identification of patients with riboflavin responsive pathologies caused by mutations in SLC25A32 that also show deficiencies in OXPHOS enzyme activities (Schiff *et al*., 2016*;* Hellebrekers *et al*., 2017).

4.2. Aims

This chapter builds on the knowledge acquired about riboflavin depletion in Chapter 3 and aims to:

- Explore the function of SLC25A32 in flavin distribution using a knockout cell line.
- Investigate the impact of severely limited flavin on complex I assembly, structure and activity.
- Characterise the wider mitochondrial and cellular changes that occur as a result of the knockout using proteomic and transcriptomics approaches.

4.3. Preliminary study: purification of the mitochondrial transporter SLC25A32

In collaboration with Dr Edmund Kunji and co-workers I attempted to characterise the transport function of SLC25A32. The yeast homologue FLX1 and human protein SLC25A32 had both been successfully cloned and overexpressed with a His-tag in *S. cerevisiae* by Dr Homa Majd. Due to the low yield of SLC25A32, the yeast homologue FLX1 was pursued instead of the human protein. FLX1 was isolated from mitochondria in the detergent 2-didecylpropane-1,3-bis-β-D-maltopyranoside by affinity chromatography and the His-tag removed; a single pure protein band was obtained, confirmed by SDS-PAGE. Recently, a technique has been developed to probe possible substrates for MCFs using the thiol-recative fluorophore *N*-[4-(7-diethylamino-4-methyl-3-courmarinyl)phenyl]maleimide (CPM) in a thermostability assay (Crichton *et al*., 2015). This is based on the rationale that, upon denaturation, protein unfolding exposes buried cysteine residues, which would react with CPM to increase the fluorescence detected, providing a melting temperature estimation for the carrier (T_m). The binding of a substrate would confer additional stability; therefore a shift in T_m would be expected. Thermostability assays with the *S. cerevisiae* ADP/ATP carrier showed that the high affinity inhibitor CATR caused a considerable shift from T_m 44.7 °C to 72.0 °C (Crichton *et al*., 2015). Therefore this approach was undertaken with FLX1, to first establish a T_m and then, by the introduction of potential substrates, determine any stability increase through binding. Unfortunately due to the inherent instability of the FLX1 during the assay, a reliable T_m could not be obtained, despite multiple attempts and protocol alterations; therefore this part of the project was not continued. Instead, work was focused upon cell lines and genetic disruption of SLC25A32.

4.4. Gene knockout using CRISPR technology to abolish target protein expression in cell lines

4.4.1. An introduction to CRISPR gene editing technology

CRISPR is based on the adaptive immune response of bacteria towards invading viruses. The process involves excising and acquiring a small fragment of phage DNA, by inserting it into a CRISPR locus, within their own genome, as a protospacer, allowing bacteria to recognise foreign DNA for degradation. Bioinformatic analysis demonstrated that protospacers were selected using adjacent downstream motifs of 3-5bp (Ran *et al*., 2013). These protospacer adjacent motifs (PAMs) are important for two of the three CRISPR system types and specific to the Cas9 expressed.

The nuclease Cas9 targets specific DNA sequences using a small guide RNA and promotes a double strand break (DSB). Two major repair mechanisms exist in bacteria to address DBSs: error prone non-homologous end-joining (NHEJ) and high-fidelity homology-direct repair (HDR). Both pathways can be utilised in gene editing by CRISPR as shown in Figure 4.3. Three types of CRISPR systems have been identified in a range of bacteria and archaea. The best understood is Type II and the system derived

from *Streptococcus pyogenes* is the basis for most CRISPR engineering technologies (Ran *et al*., 2013). It involves Cas9 and an array of guide RNAs, which direct the nuclease towards a 20bp target at a specific locus. For target sequences to be recognised and CRISPR edited in *S. pyogenes,* the PAM sequence following them, at the 3' end, must be NGG. The use of CRISPR as a genetic engineering tool has many benefits, as it is a highly flexible, simple system that can be manipulated to generate knockouts, insertions and the addition of tags to genes within mammalian and other cells, despite being a bacterial system in origin.

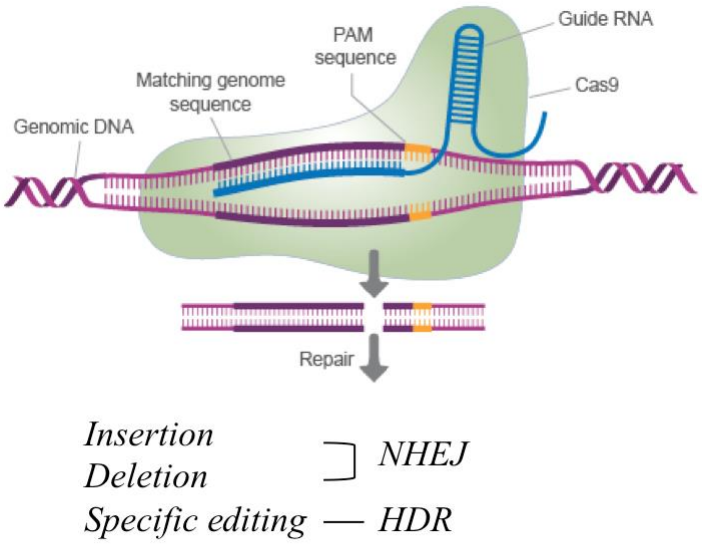


Figure 4.3 Genome editing with CRISPR. The genomic DNA in pink is targeted by a specific guide RNA (blue) requiring the complementary sequence and a 3' PAM (yellow). Cas9 (green) is directed by the guide RNA and makes a double stranded break. This may be repaired by either NHEJ or HDR. NHEJ is the most active pathway but also more error prone – therefore likely to introduce indels into the repair site. This causes the target gene to be interrupted and a frame shift may also introduce a premature stop codon. For a specific change to be facilitated a repair template must be present, to allow a point mutation or larger insertion (such as a tag) to be included in the repair. However this is a much less efficient pathway and some NHEJ will also take place. Adapted from http://www.transomic.com/Products/CRISPR-Genome-Editing.aspx.

To investigate the function of SLC25A32 a similar approach to that explored in Chapter 3 was employed. By disrupting the abundance of this transporter it would be possible not only to determine the effect on complex I and other flavin-reliant mitochondrial proteins but also demonstrate its role in transport of flavins across the IMM. To date, despite the work of Spaan and co-workers to address the similarities between yeast FLX1 and SLC25A32 (sharing 29% sequence identity and 44% similarity), there has been no work on the human transporter in its native system (Spann *et al*., 2005). Instead, it has been used to rescue the function of mutants in non-human model organisms. Therefore investigations in a human cell line would be informative and add breadth to the existing work on this protein. Knocking out SLC25A32 by CRISPR would generate a cell line without the functioning transporter, which could be characterised in detail to help understand the effects of a non-functioning transporter and attempt to further uncover its physiological role. This approach would overcome the inherent problems of studying membrane proteins by overexpression and purification and provide a stable system that could be tested by multiple complementary techniques to build an understanding of SLC25A32 function. Additionally,

it would immediately demonstrate whether a null mutation was lethal or if survival was still permitted despite a loss of function frame shift.

4.4.2. SLC25A32 knockout in HAP1 cells

CRISPR allows the targeted knockout of genes by the introduction of indels, but this must be achieved for each gene copy. Therefore the higher the copy number the more difficult it is to achieve a full knockout. 143Bs are known to display genomic instability and also contain more than two copies of most chromosomes. SLC25A32 is found on chromosomes 8. Karyotyping was undertaken by Cell Guidance Systems and the five 143B cells tested all displayed a range of copy numbers; three or more cells contained an additional copy of chromosomes 3, 4, 5, 10, 12, 13 and 16, with two extra copies of chromosome 8. This highlighted that 143Bs were not the most suitable choice for CRISPR knockout and so another cell line was chosen, the adherent human KBM-7 derived cell line HAP1. These cells are haploid, which therefore simplified the knockout process.

A commercially available SLC25A32 knockout (referred to as SLC25A32 KO) was obtained from Horizon Discovery, along with the parental HAP1 line. The deletion in exon 2 was confirmed by sequencing, Figure 4.4, and these cells were taken forward for conditioning and investigation into the effects on complex I. Additionally, I used CRISPR to design guide RNAs targeting exon 1 of the SLC25A32 gene, to generate a number of additional knockouts in HAP1 (Chapter 2, section 2.3.3), with a view to validating results obtained using the commercially available SLC25A32 KO (results are discussed in section 4.9). All SLC25A32 knockout cells and HAP1 controls were conditioned (riboflavin-free DMEM + 10% FBS for one month) prior to further investigation.

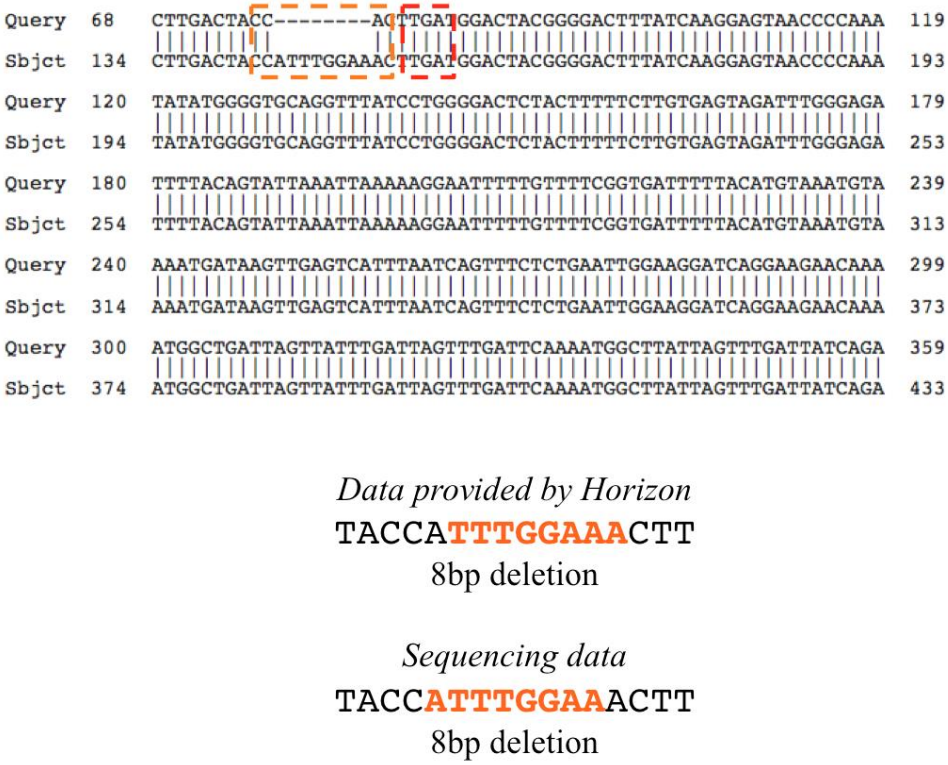


Figure 4.4 Sequencing data alignment for SLC25A32 knockout. Sequencing confirmed an 8bp deletion (orange) in exon 2, although differing by one base pair compared to data provided by the supplier (Horizon). This brings a stop codon into frame and therefore would be expected to produce a truncated protein (<80 amino acids) or cause exon skipping, to produce a non-functional protein.

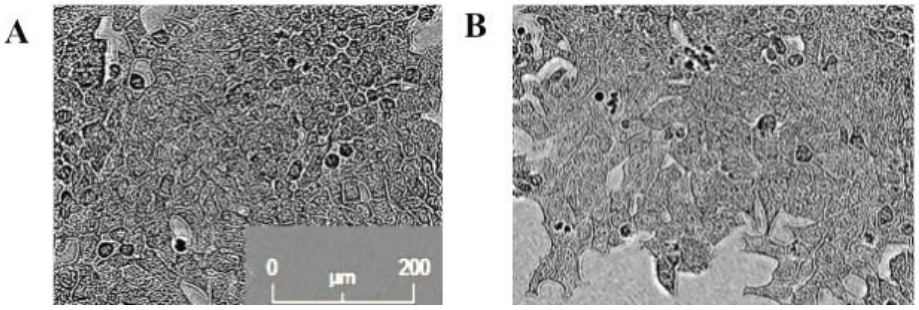


Figure 4.5 HAP1 and SLC25A32 knockout cells grown under high flavin conditions in IMDM. Cells were seeded at 14,000/well with 4 ml of IMDM + 10% FBS and incubated at 37 ºC for 160 hours. Incucyte software was used to record the change in growth. **A.** HAP1 cells showed normal fibroblast-like morphology, <50 µm in diameter. **B.** SLC25A32 KO also showed fibroblast-like morphology, with some elongations observed.

Figure 4.5 demonstrates the morphology of the SLC25A32 KO and HAP1 parental cell lines grown in IMDM (as recommended by the supplier). On conditioning the SLC25A32 KO (SLCKO) showed a change in morphology and growth rate, which was progressive and correlated with lower riboflavin availability. Figure 4.6 shows the considerable difference in growth profiles of the two cell lines undergoing conditioning, in riboflavin-free DMEM; SLCKO cells grew much more slowly and after more than 300 hours were not confluent. This was accompanied by a change in morphology shown in Figure 4.7B. Cells became increasingly elongated with fewer cell-to-cell contacts. Consequently these initial observations indicated that riboflavin depletion resulted in changes to the SLCKO cells compared to the HAP1 controls. Further work then explored these differences in the context of complex I structure and function.

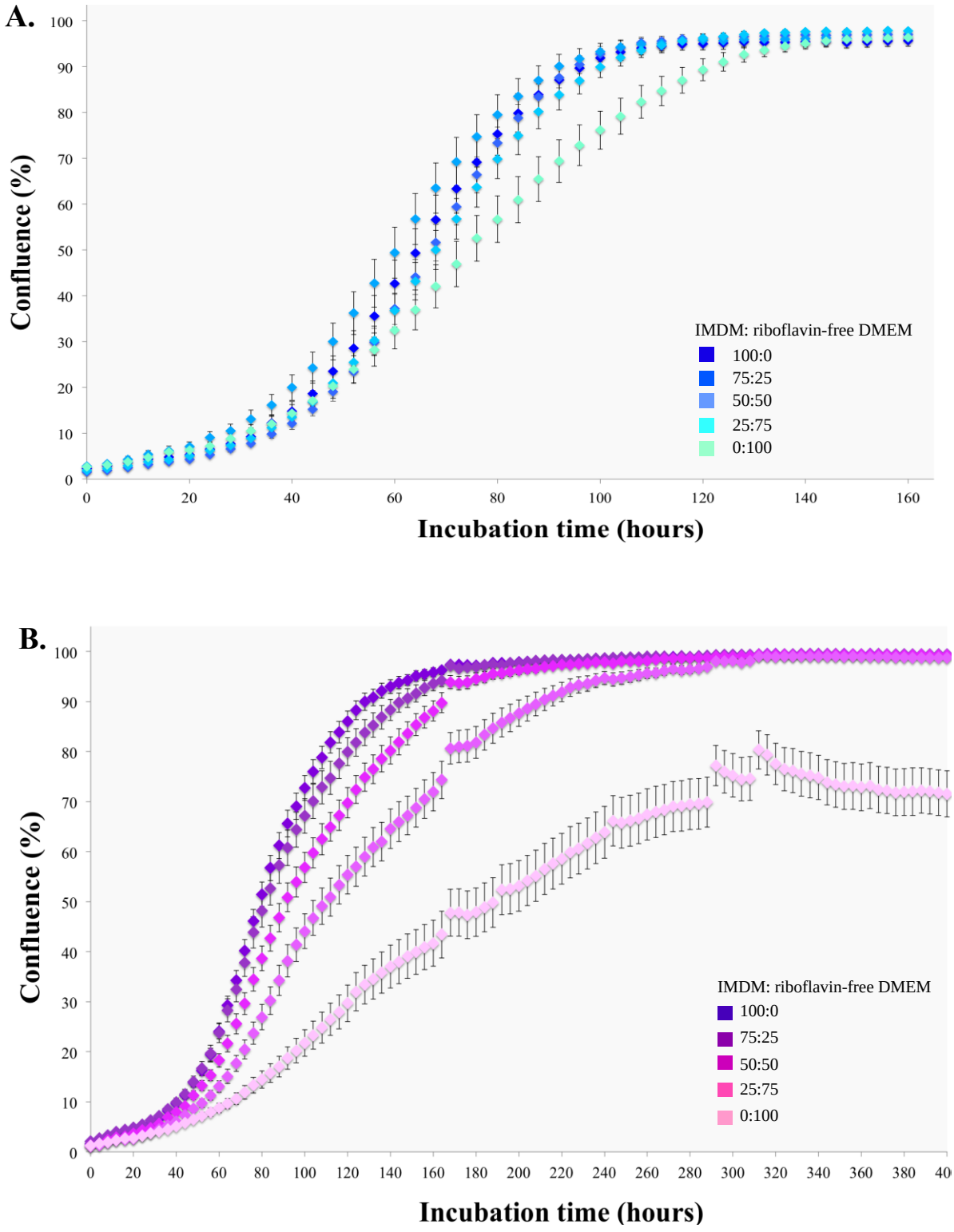


Figure 4.6 Growth profile of HAP1 during conditioning stages. Cells were seeded at 14,000/well with 4 ml of appropriate medium and incubated at 37 ºC for at least 160 hours. Incucyte software was used to record the change in growth, with images obtained every 4 hours and percentage confluence calculated. Error bars denote standard error. Cells were adapted from IMDM to riboflavin free DMEM in 25% increments after two passages. **A.** HAP1 parental line – IMDM (dark blue), three increasingly lighter shades of blue for IMDM: riboflavin free DMEM 75:25, 50:50, 25:75 and riboflavin-free DMEM (mint green). **B.** SLC25A32 knockout – IMDM (purple), three increasingly lighter shades of purple for IMDM: riboflavin free DMEM 75:25, 50:50, 25:75 and riboflavin-free DMEM (light pink). These cells were monitored for 400 hours to determine whether confluence would be reached, however the erratic nature of the later data points demonstrates the increasing levels of cell death in the riboflavin free DMEM knockout cell wells. Data from one representative assay are shown.

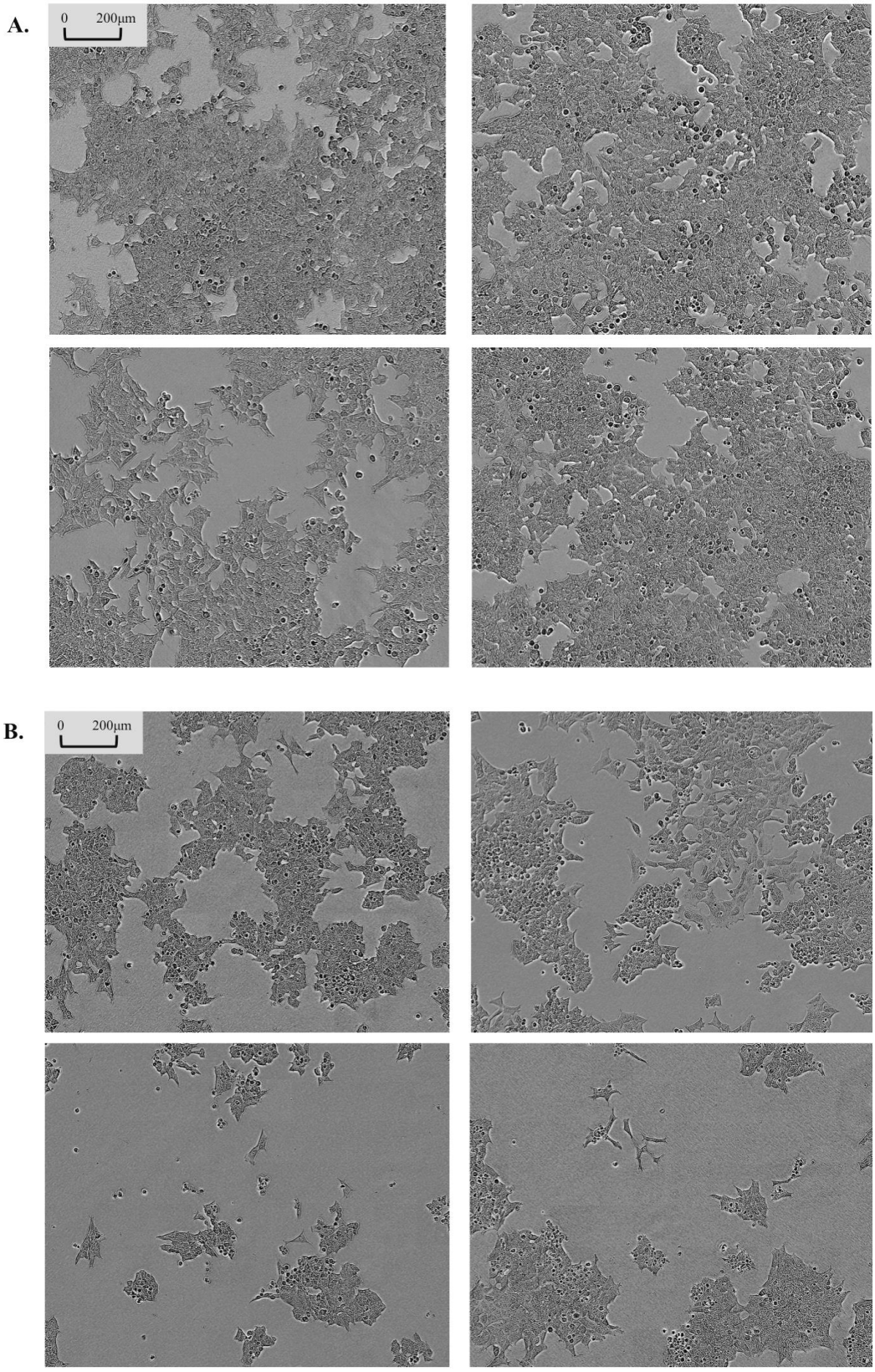


Figure 4.7 Time interval cell growth images to compare conditioning of HAP1 and SLC25A32 cells. Cells were seeded at 14,000/well with 4 ml of appropriate medium and incubated at 37 ºC for 160 hours. Incucyte software was used to image the wells in 16 segments every 4 hours. **A.** HAP1 conditioning – top panels (L-R) IMDM: riboflavin free DMEM 75:25 and 50:50, bottom panels (L-R) 25:75 and riboflavin-free DMEM. **B.** SLC25A32 knockout conditioning – top panels (L-R) IMDM: riboflavin free DMEM 75:25 and 50:50, bottom panels (L-R) 25:75 and riboflavin-free DMEM.

4.5. Characterisation of SLC25A32 knockout in riboflavin depleted HAP1s

4.5.1. Respiratory phenotype

To determine whether the SLC25A32 knockout differed in respiratory profile compared to the control under low riboflavin conditions a Seahorse X96 analyser was used to monitor the OCR and ECAR during a stress test (as described in Chapter 2, section 2.6.1). Figure 4.8 shows the striking and significant difference in oxygen consumption between the two cell lines: basal OCR of the SLC25A32 KO was on average only 11% of the control. Also there was no observable response to the additions of oligomycin (after 40 minutes), FCCP (after 80 minutes) and rotenone & antimycin A (after 130 minutes), whereas the uncoupled OCR of the control increased by 50% on addition of FCCP, compared to its basal rate. This suggests that the knockout cells showed predominantly non-mitochondrial respiration. Interestingly no difference in ECAR was observed between the SLC25A32 KO and control (data not shown) suggesting that the knockout cells had not shifted towards glycolysis to meet energy demands.

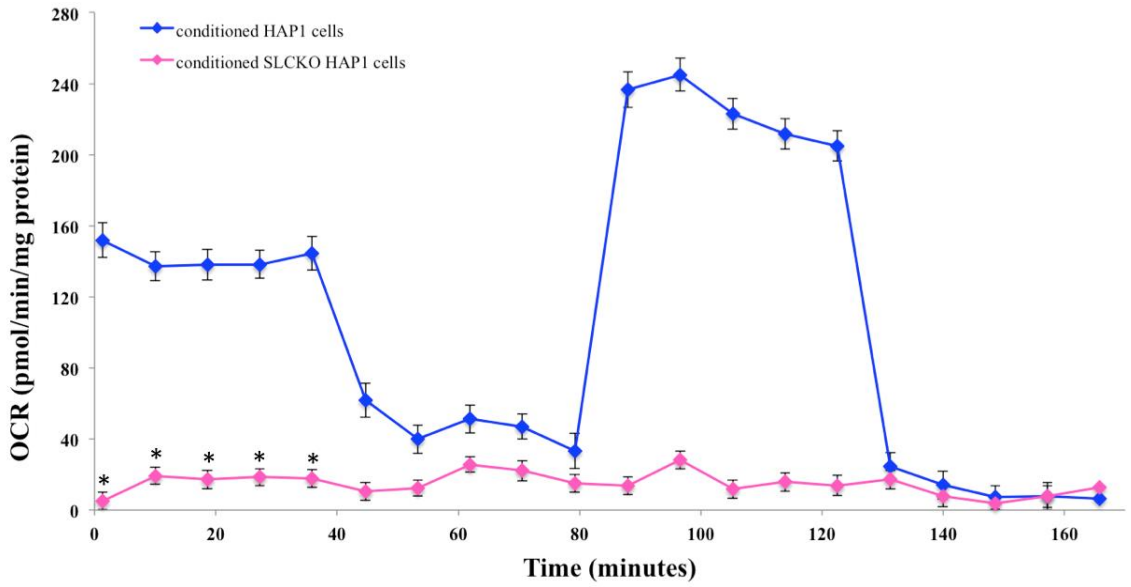


Figure 4.8 Seahorse extracellular flux data comparing the respiratory profile of conditioned SLC25A32 knockout HAP1 cells with the parental line. The control HAP1 cell line is shown in blue and the SLC25A32 KO in pink. Both cell lines were conditioned for one month and seeded at 11,500 per well the day before the assay. 46 wells of each condition were assayed and the experiment repeated thrice (mean results for one representative assay shown, with error bars denoting standard error). OCR was corrected for seeding density using BCA protein quantification of wells after the assay. Significantly different basal OCR measurements between the two cell lines are denoted by * ($P<0.0001$ by unpaired T-test).

4.5.2. Mitochondrial membrane potential of SLC25A32 knockouts

The OCR data suggested that the OXPHOS system is inactive and therefore the mitochondrial membrane potential was investigated, as a non-functional respiratory chain is expected to result in low

membrane potential. The fluorescent dye TMRM is a cell permeant cation that is actively accumulated by mitochondria and can be used to indicate the polarisation status of a cell population (Scaduto & Grotyohann, 1999). Uptake was measured after 30 minutes incubation and HAP1 samples were used as a control to set the gating threshold, distinguishing polarised and non-polarised cells. This gate was then applied to all samples for comparison between the knockout and parental cell lines. Figure 4.9 shows that, surprisingly, the knockout cells showed an average polarised population of 88% compared to 90% of the parental HAP1 cells. Additionally, FCCP uncoupled both cell types to a similar extent, suggesting that SLC25A32 maintains a membrane potential despite the OCR data demonstrating a lack of oxidative phosphorylation.

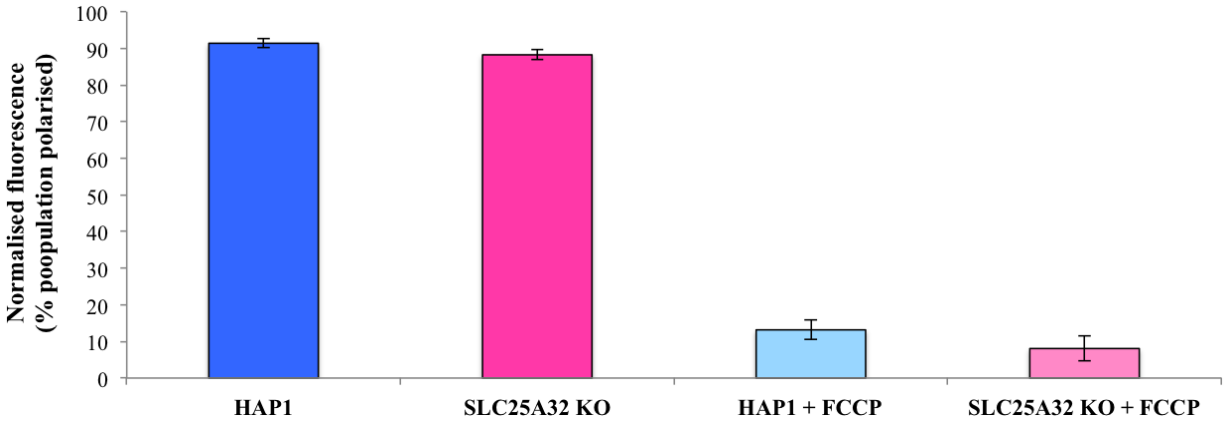


Figure 4.9 Proportion of cell population polarised using TMRM fluorescence. Triplicate cell samples were treated with TMRM for each cell line and the dye retention measured by fluorescence. 5 μM FCCP was added and after incubation samples were tested again for TMRM fluorescence. Data were normalised using Hoerchst DNA staining measurements. Gating was used to determine the polarised proportions of each population, using the HAP1 controls, and overlaid on each plot. Error bars denote standard deviation.

To maintain a membrane potential without oxidative phosphorylation indicated that SLCKO cells employed another method, comparable to that employed by mtDNA-lacking ρ^0 cells. In ρ^0 cells the ATP:ADP transporter acts in reverse to import ATP, which is then broken down by ATP synthase also running backwards. ADP and P_i are then exchanged out of the mitochondria, via the phosphate carrier protein (SLC25A3) (Buchet & Godinot, 1998). This mechanism may provide an explanation for the results obtained from SLCKO cells regarding their low rotenone/antimycin insensitive oxygen consumption, despite presence of a membrane potential. However these results do not detail which components of the respiratory chain are most affected by the knockout or the impact upon complex I.

4.5.3. OXPHOS chain activity

To probe which of the individual complexes was most affected, in-gel activity assays were used to assess the activity of the flavin containing complexes I and II and also complex IV.

4.5.3.1. Complex I in-gel activity assay

Complex I activity was assessed using the electron acceptor NBT, which on reduction produces a purple band of activity on BN-PAGE gels of purified mitochondria (Chapter 2, section 2.4.2). Figure 4.10 shows in-gel complex I activities for both DDM and digitonin solubilised mitochondria from conditioned HAP1 and SLC25A32 KO cells.

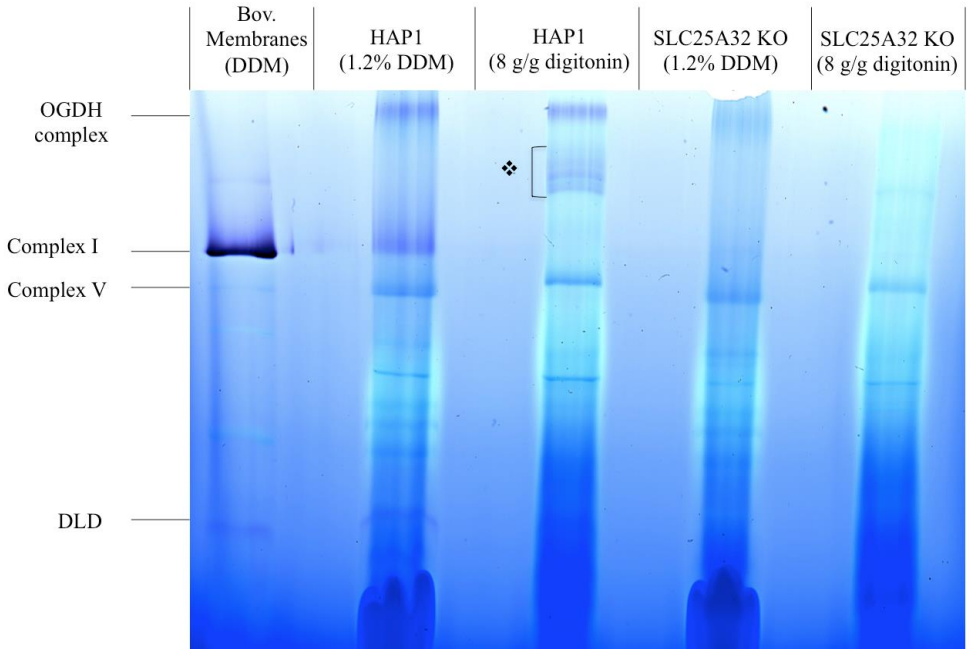


Figure 4.10 Complex I in-gel activity characterisation of the SLC25A32 knockout. Purified mitochondria were solubilised with 1.2% DDM or 8g/g digitonin. Complex I activity assay was carried out using 0.5 mg/ml NBT, 100 µM NADH in 20 mM Tris pH 7.4.

Purple bands present in Figure 4.10 indicate protein with flavin site activity; NADH is oxidised by flavin and passed to the NBT electron acceptor. Therefore, other flavin active proteins were also highlighted. Oxoglutarate dehydrogenase complex (OGDH) contains the E3 component dihydrolipoyl dehydrogenase (DLD) with a FAD cofactor. This large complex is usually present ~2,500 kDa and on DDM solubilisation the DLD component is also observed at lower molecular weight ~ 55 kDa representing its monomeric state (Wittig *et al*., 2006). Therefore three bands were visible in the HAP1 DDM solubilised sample, an active complex I band corresponding to the same position as found in the bovine mitochondrial membrane standard, along with DLD in OGDH and as a monomer. Digitonin acts as a milder detergent and so larger complexes remain intact; therefore DLD was only present in the large OGDH band. A series of high molecular weight bands with flavin activity demonstrated different complex I containing supercomplexes (❖). Complex I is found tightly associated with complexes II, III and IV in differing stoichiometries in the IMM and solubilisation with digitonin preserves these large assemblies. Whilst the HAP1 samples showed the active bands expected the SLCKO samples lacked all

flavin activity; therefore suggesting that complex I in these cells was not active and that the knockout had broad effects on other mitochondrial flavoproteins, as evidenced by lack of active DLD. These observations were not a result of differential mitochondrial protein loading due to the similar levels of ATP synthase present. This band ~ 700 kDa is usually visible after BN-PAGE even without staining and was clearly present in all four cell samples, suggesting loading was consistent; although protein migration differed slightly between DDM and digitonin solubilised samples.

4.5.3.2. Complex II in-gel activity

Complex II in-gel activity assays revealed a distinct difference in the HAP1 control compared to the SLC25A32 knockout. In Figure 4.7 two loading volumes were used for each cell line - the complex II activity in the HAP1 samples increased proportionately, as did the intensity of the complex V band. However in the SLC25A32 KO samples there was no visible complex II band, even with the high loading volume. Therefore, from Figure 4.10 and 4.11, it was apparent that the KO cell line lacked functioning complex I and II, possibly due to these complexes lacking their flavin cofactor.

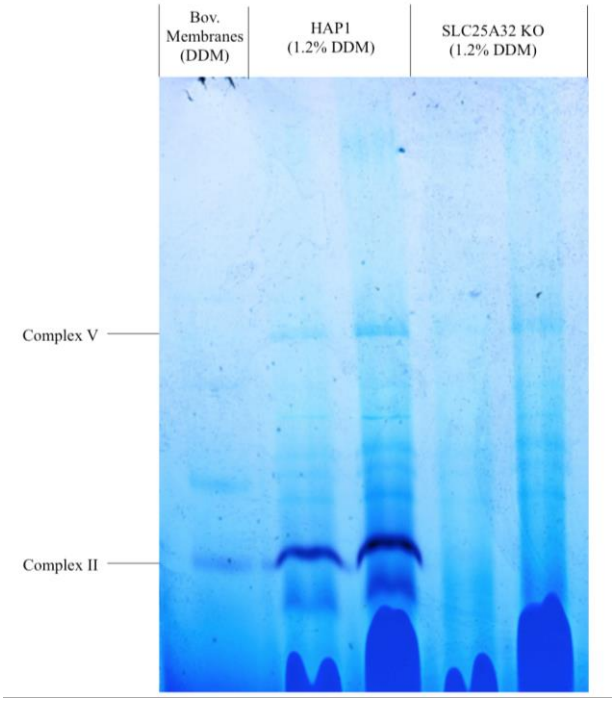


Figure 4.11 Complex II in-gel activity characterisation of the SLC25A32 knockout. Purified mitochondria were solubilised with 1.2% DDM. Complex II activity assay was carried out using 84 mM succinate, 2mg/ml NBT, 10 mM sodium cyanide and 0.2 mM phenazine methosulphate (PMS) in 20 mM Tris pH 7.5.

4.5.2.3. Complex IV in-gel activity

To complete the in-gel characterisation of SLC25A32 KO mitochondria a complex IV assay was also undertaken. The complex III assay, using diaminobenzidine (DAB), gave unreliable results, even for the bovine heart membrane control, consistent with discussion in the literature (Jung *et al*., 2000; Smet *et al*., 2011). In Figure 4.12 two loading volumes were used for each cell line and in both samples complex IV activity was apparent, increasing proportionately with sample loading. The position of the complex IV active band in the cell line samples was equivalent to that of the bovine mitochondrial membrane standard. Therefore, whilst the activities of complex I and II were compromised in the knockout complex IV was unaffected.

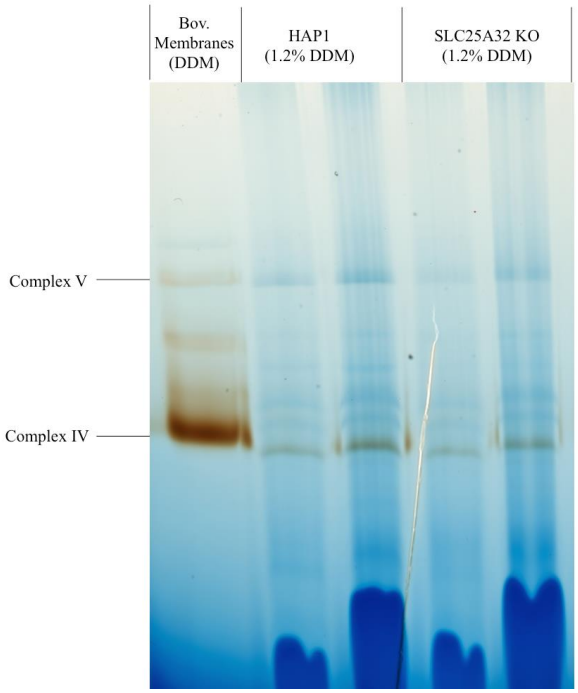


Figure 4.12 Complex IV in-gel activity characterisation of the SLC25A32 knockout. Purified mitochondria were solubilised with 1.2% DDM. Complex IV activity assay was carried out using 0.4 mg/ml 3-3 diaminobenzidine (DAB) and 1 mg/ml cytochrome c in 50 mM phosphate, pH 7.2.

4.6. The effect of SLC25A32 knockout on the mitochondrial proteome

4.6.1. Changes in protein abundance were observed in SLC25A32 knockouts by Western Blot

For an initial investigation into differences in protein abundance between the SLC25A32 KO and parental HAP1s, proteins were chosen that had shown decreases in previous work with the RFK knockdown 143B cells. Western Blot using fluorescent secondary antibodies allowed the extent of change to be quantified.

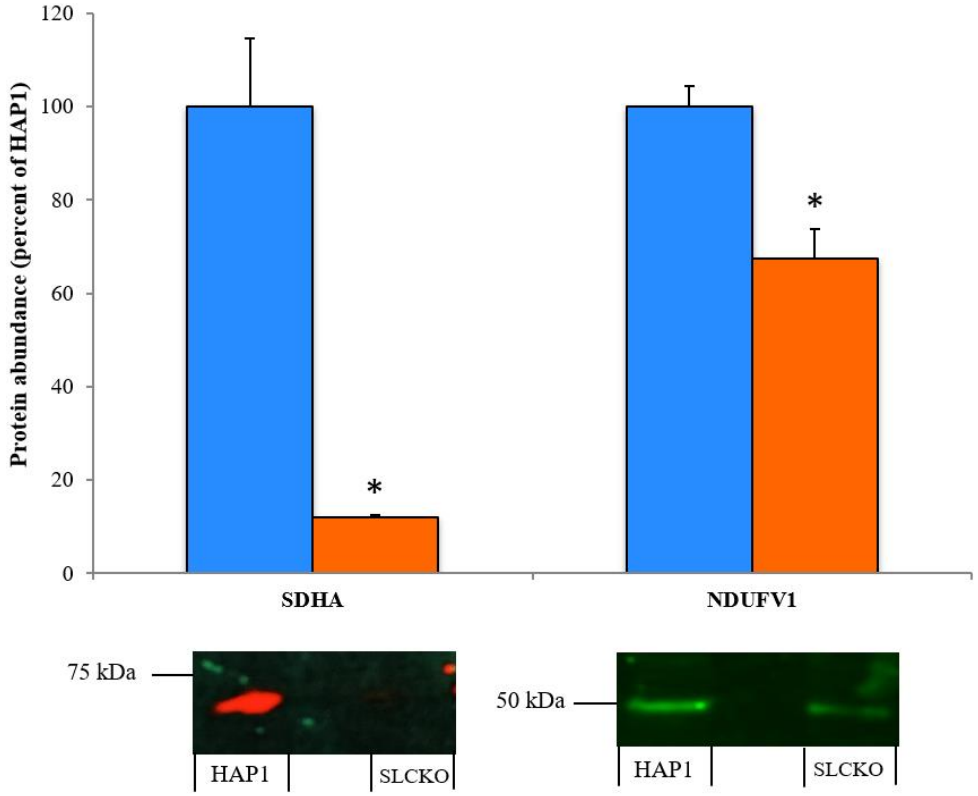


Figure 4.13 Quantitative Western Blot for flavin containing subunits of complex I and II. 20 μg cell lysates from conditioned HAP1s and SLCKO were run on SDS-PAGE and probed for protein abundance. Loading was accounted for by the signal from VDAC1 or Drp1 antibodies and data adjusted. Protein abundance in SLC25A32 KO is shown as a percentage of HAP1 protein levels. A representative blot for SDHA and NDUFV1 are shown. Error bars denote standard error and * show a significant decrease, $P<0.05$, by unpaired T-test.

Figure 4.13 shows data for NDUFV1 and SDHA, proteins which showed large decreases in previous data, two-fold and four-fold respectively. In the SLC25A32 knockout these proteins also showed differing abundance, <20% for SDHA and <70% for NDUFV1 compared to the control. Due to their roles as the flavin cofactor binding subunits in complexes I and II this correlates with the activity data from Figures 4.6 and 4.7, demonstrating inactive complex I and II. These initial experiments indicated that SLCKO displayed considerable differences in key complex I and II subunits and this could extend to other mitoflavoproteins; therefore a more comprehensive review of these changes by SILAC was undertaken.

4.6.2. SILAC confirmed changes in protein abundance when SLC25A32 was knocked out in conditioned HAP1s

SILAC was performed on conditioned HAP1 and SLC25A32 knockout cells to compare relative protein abundances. Ideally the experiment would have been repeated three times in each labelling orientation to provide robust data. However both cell lines displayed limited growth rates and viability in SILAC conditions, independent of labelling orientation, and so data could only be collected once for each labelling direction. This effect may be due to the necessary use of dialysed FBS (Chapter 2, section 2.7.1). From the SILAC analysis 2805 proteins were detected in both cell lines, across both labelling orientations, to provide a heavy:light labelled (H/L) ratio. Figure 4.10 displays the data as Log_2 H/L, where the SLC25A32 knockout was heavy-labelled and HAP1 un-labelled in the forward direction. Therefore, proteins with positive values were increased in the knockout and those with negative values decreased in abundance. The reverse labelling ratios were inverted to account for the opposite labelling - so that positive and negative values were comparable to the forward orientation. Those proteins with a three-fold or greater change in abundance in both labelling orientations were deemed considerably different between the samples (Log_2 ratio $\geq$ 1.58) and those with a greater than two-fold change highly changed (Log_2 ratio $\geq$ 1.00).

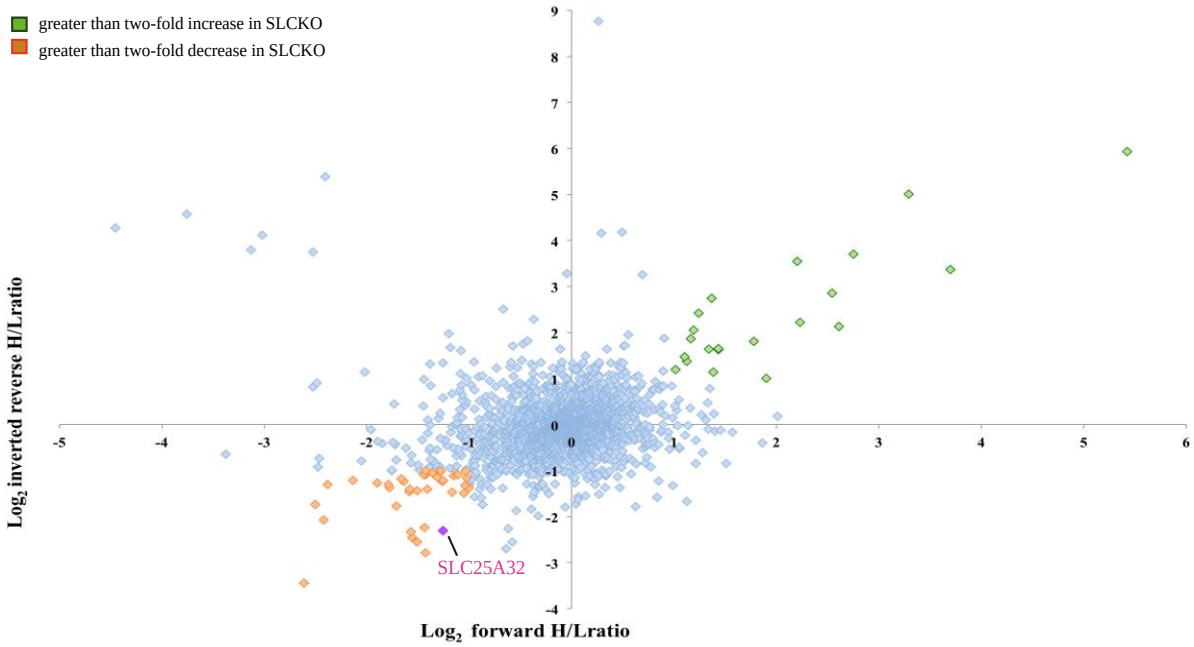


Figure 4.14 SILAC data comparing protein abundance from conditioned HAP1 and SLC25A32 KO cells. Single repeat data from each labelling orientation are shown. Data are displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio; such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the SLC25A32 KO and negative values indicate decreased abundance in the knockout. Proteins with greater than two-fold increase (in both labelling orientations) are highlighted in green and those with a greater than two-fold decrease in the knockout are in orange. SLC25A32 protein abundance is shown in pink.

Figure 4.14 highlights that most proteins clustered around the axes intersection, showing unchanged abundance ($Log_2 = 0$), whereas 38 were more than two-fold decreased in the knockout, highlighted in orange, and 21 were increased, shown in green (Appendix 7.6). Of the proteins with decreased abundance in the knockout 17 were non-mitochondrial, two were predicted mitochondrial and 19 were mitochondrial proteins. These included six complex III subunits, five mitoflavoproteins, one ATP synthase subunit and one complex II subunit. Of the non-mitochondrial proteins the plasma membrane riboflavin transporter SLC52A2 was more than three-fold lower than the control; however this must be interpreted with caution as the sample should not contain non-mitochondrial elements and so these proteins may not be equally represented by both samples.

The SLCKO contains a deletion in exon 2 of SLC25A32, yet the protein was detected in both cell lines - though with a greater than two-fold decrease in the knockout. Seven unique peptides were detected by SILAC, with sequence coverage of 29.2%. Figure 4.15 shows the position of these peptides along the protein sequence; the six peptides (grey) detected in both cell lines covered the length of the protein but did not include sequence from exon 2 (underlined), where the 8 bp deletion is found (highlighted in red). The peptide missing from SLC25A32 KO cells (yellow) crosses the exon boundary from exon 1-2 and indicates that exon 2 was only translated in the HAP1 cells. Therefore, in the knockout, peptides could be detected from non-functional protein generated through alternative splicing to skip exon 2 – with protein generated from exons 1 and 3-7. Peptides corresponding to sequence from exon 2 were only detected in the HAP1 cells, as all seven exons were translated. Exon skipping induced by CRISPR deletions has been reported previously and allows the translation of protein despite a disrupted reading frame – usually resulting in a non-functional protein, although it also has the potential to encode gain of function or partially functional proteins (Mou *et al*., 2017).

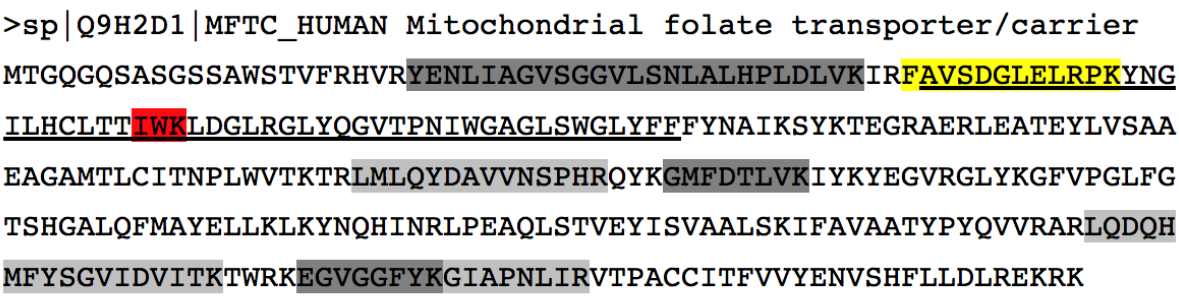


Figure 4.15 SLC25A32 peptides identified by mass spectrometry during SILAC analysis comparing protein abundance from conditioned HAP1 and SLC25A32 KO cells. The protein sequence for SLC25A32 is annotated with the seven unique peptides detected by mass spectrometry in the SILAC experiments. Those in grey were found in both the HAP1 and SLC25A32 cell lines – different shades are used because the final peptides (starting EGVG and GIAP) cover overlapping sequence (both include the sequence in light grey) but represent two unique peptides rather than one. The yellow peptide was only found in HAP1 cells. Exon 2 is underlined and the three residues disrupted by the 8 bp deletion in the knockout are highlighted in red.

4.6.2.1. N-domain complex I subunits were the most decreased in abundance across the whole enzyme in conditioned SLC25A32 knockout cells

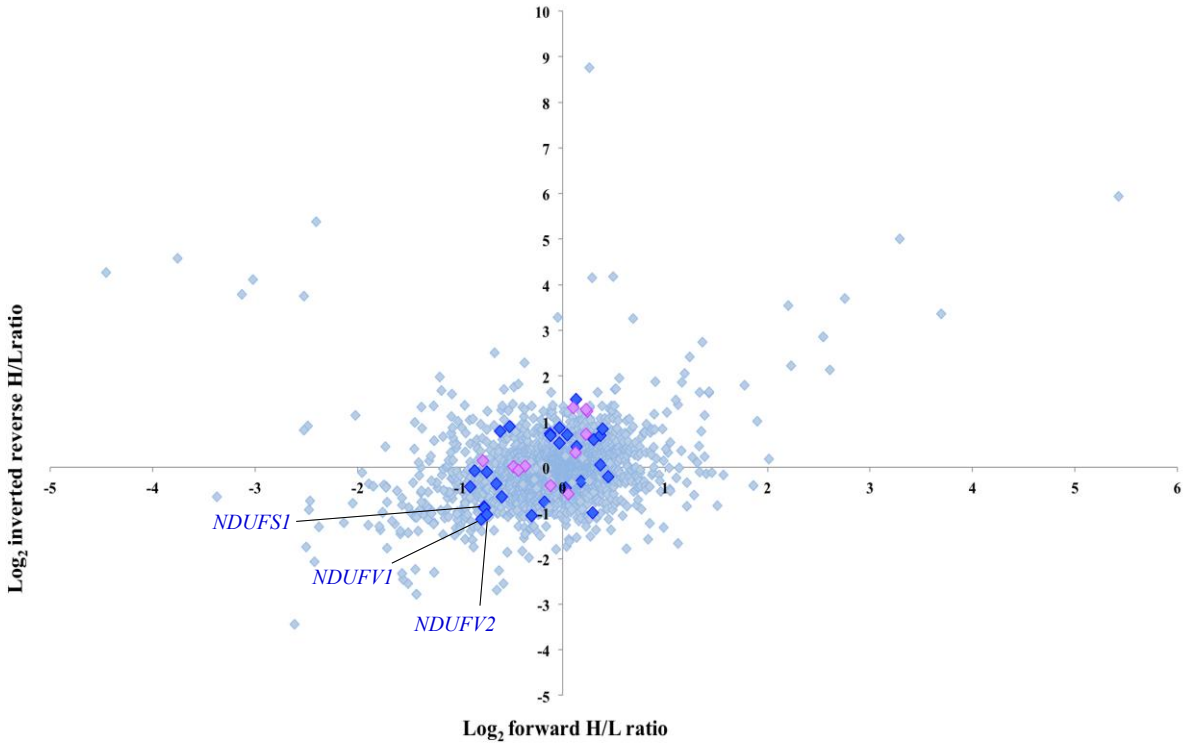


Figure 4.16 SILAC data comparing protein abundance from conditioned HAP1 and SLC25A32 KO cells. Single repeat data from each labelling orientation are shown. Data are displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio; such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the SLC25A32 KO and negative values indicate decreased abundance in the knockout. Complex I subunits are highlighted in dark blue and assembly factors in purple.

Complex I subunit coverage in each orientation was over 80%, however only 28 subunits were detected in both directions to provide H/L ratios (Appendix 7.6). As shown in Figure 4.16, none displayed a greater than two-fold change in abundance, although NDUFV1, NDUFV2 and NDUFS1 were among the most decreased subunits, with a decrease in abundance of just below two-fold. Additionally the 11 assembly factors detected displayed unchanged abundance when compared to the HAP1 control. It is interesting that despite the very low OCR of the knockout and lack of in-gel activity for complex I, there was no striking difference in abundance of complex I subunits. These differences are in line with the moderate decrease in NDUFV1 of 33%, as detected through Western Blot. However neither SILAC nor Western blotting give any indication of complex I structure, therefore whilst the abundance of complex I subunits may be largely unchanged the proportion present as the fully assembled enzyme may differ.

4.6.2.2. Complex III subunits were highly decreased in the conditioned SLC25A32 knockout cells

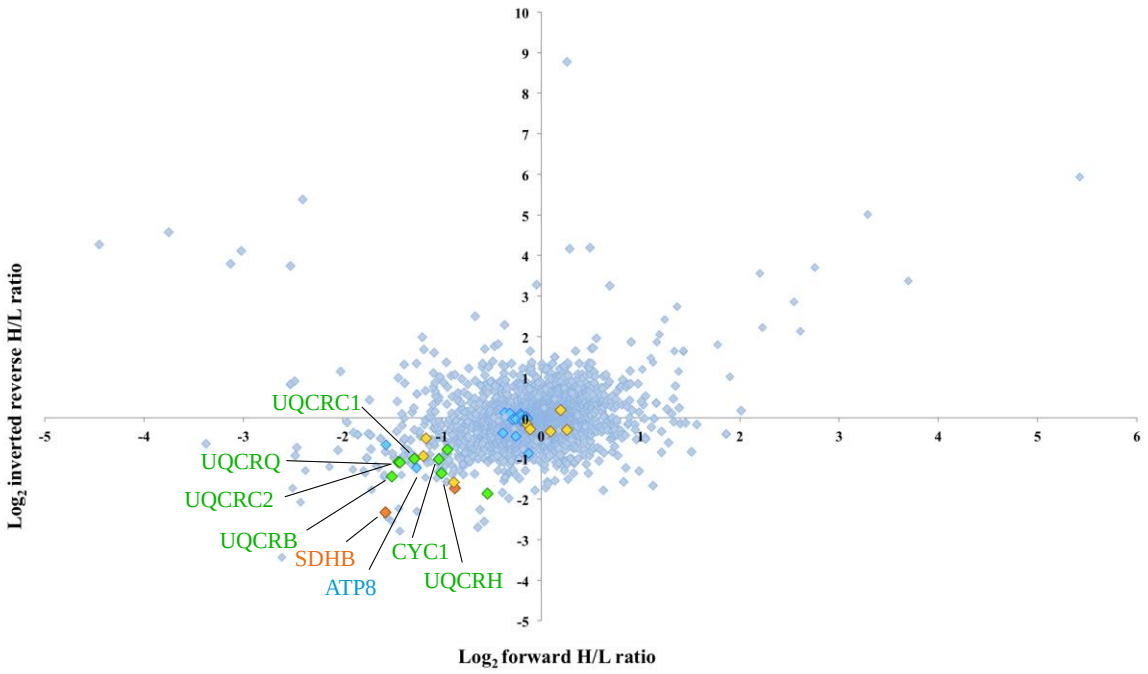


Figure 4.17 SILAC data comparing protein abundance from conditioned HAP1 and SLC25A32 KO cells. Single repeat data from each labelling orientation are shown. Data are displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio; such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the SLC25A32 KO and negative values indicate decreased abundance in the knockout. Complex II subunits are shown in orange, complex III subunits in green, complex IV in gold and complex V subunits are highlighted in light blue. Subunits with a greater than two-fold decrease in the conditioned knockout (in both labelling orientations) are labelled.

The abundance of subunits from the remaining respiratory chain complexes was also assessed. Figure 4.17 highlights all subunits detected: two from complex II (orange), eight complex III subunits (green), eight from complex IV (gold) and 15 from complex V (cyan) (raw data in Appendix 7.6). These subunits were either unchanged in abundance or decreased in the knockout. Complex II subunit SDHB showed decreased abundance of more than two-fold and SDHA was also less abundant, decreased by at least 1.8-fold. This difference in subunit levels correlates with the lack of complex II in-gel activity in the knockdown (Figure 4.11) and lowered abundance detected by Western Blot for SDHA (although this was more pronounced at ~ five-fold, Figure 4.13). FAD is covalently bound to SDHA, unlike non-covalent association of FMN with NDUFV1, and this may confer some stability; therefore lack of the cofactor may have a larger effect on protein structure (Kim *et al*., 2012).

In contrast complex V subunits were largely unchanged, with the exception of ATP-8 that was lowered by over two-fold. However as a small hydrophobic subunit this may be an artefact of mass spectrometry and not accurately represent protein abundance. Complex IV subunits were also mostly unchanged, although the zinc binding COX5B and dimerising subunit COX6B1 were less abundant. This could

indicate some differences in turnover of complex IV subunits in the knockout, although Figure 4.12 demonstrates that complex IV activity was comparable to the control.

The most striking change in respiratory complexes relates to complex III. Figure 4.13 demonstrates a shift in all eight detected subunits towards a decrease in the knockout and the abundance of six subunits was more than two-fold lower. Although this complex does not contain a flavin cofactor, it is immediately downstream of complex I and II, deriving electrons from ubiquinol to reduce cytochrome c. The data therefore suggest that changes to complex I and II due to flavin depletion cause changes further along the respiratory chain. The disparity lies in the large changes observed for complex III compared to the relatively unchanged protein abundances for complex IV and V subunits, as it might be expected that all downstream complexes would be affected.

4.6.2.3. A subset of mitoflavoproteins showed decreased abundance in SLC25A32 knockout cells

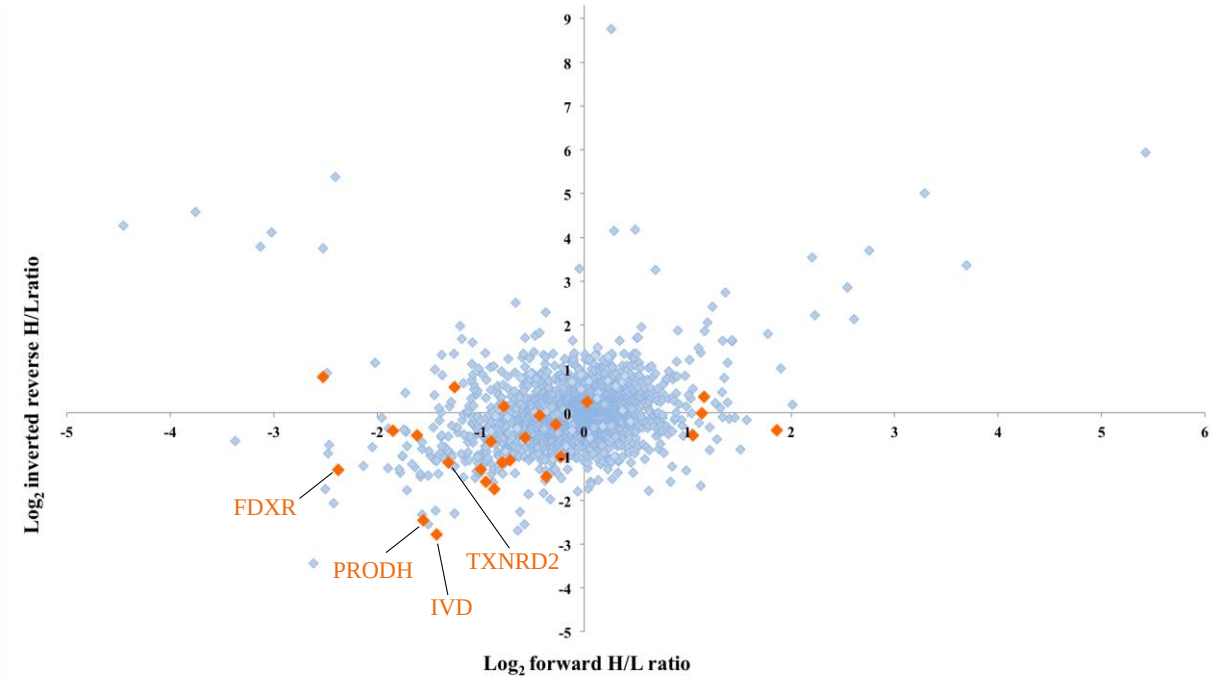


Figure 4.18 SILAC data comparing protein abundance from conditioned HAP1 and SLC25A32 KO cells. Single repeat data from each labelling orientation are shown. Data are displayed as Log_2 H/L forward ratio against the inverted Log_2 H/L reverse ratio; such that positive H/L ratios in each orientation (top right quadrant) represent proteins with higher abundance in the SLC25A32 KO and negative values indicate decreased abundance in the knockout. Mitoflavoproteins are shown in orange and those with a greater than two-fold decrease in the conditioned knockout (in both labelling orientations) are labelled.

The response of mitoflavoproteins to SLC25A32 knockout was analysed, particularly in light of a possible hierarchy observed from SILAC data in Chapter 3, section 3.10.3. Of the 32 human mitoflavoproteins 25 were detected, highlighted in orange in Figure 4.18 (raw data in Appendix 7.6).

Overall there was a shift towards lowered abundance in the SLC25A32 knockout, but as noted before there was a range of effects, not all mitoflavoproteins are affected to the same extent. Four demonstrated a greater than two-fold decrease: NADPH:adrenodoxin oxidoreductase (FDXR), isovaleryl-CoA dehydrogenase (IVD), proline dehydrogenase 1 (PRODH) and thioredoxin reductase 2 (TXNRD2). IVD and PRODH are involved in amino acid metabolism, whereas TXNRD2 is a thioredoxin reductase and FDXR is the first electron acceptor of mitochondrial P450 systems; however all require a non-covalently bound FAD cofactor (Tiffany *et al*., 1997; Bender *et al.*, 2005; Sibbing *et al*., 2011, Ziegler *et al*., 1999). Another six flavoproteins demonstrated a lower abundance in the SLCKO, including both flavoproteins from complex I and II. The acyl-CoA dehydrogenase ACADVL, the electron-transfer flavoprotein ETFDH and glutaryl-coA dehydrogenase (GCDH), which transfers electrons to ETFDH, were also decreased, along with the porphyrin-related oxidase PPOX. Five other mitoflavoproteins also showed depletion, however the extent of this change was not consistent between the labelling orientations.

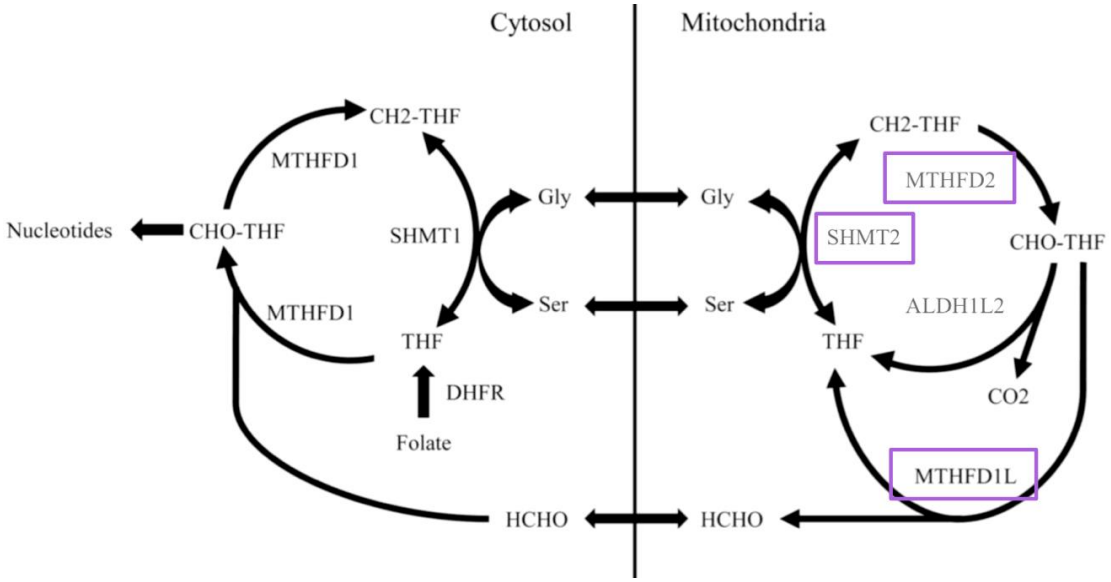


Figure 4.19 Schematic representation of folate metabolism and the compartmentalisation into the cytosol and mitochondria. Enzymes responsible for catalysis at each step are written next to the arrows. Cytosolic enzymes: MTHFD1, methylenetetrahydrofolate dehydrogenase 1, SHMT1, serine hydroxymethyltransferase, DHFR, dihydrofolate reductase. Mitochondrial enzymes: ALDH1L2, aldehyde dehydrogenase 1 family member L2, MTHFD1L, methylenetetrahydrofolate dehydrogenase 1-like, MTHFD2, methylenetetrahydrofolate dehydrogenase 2 – those detected in SILAC are highlighted in purple. Figure adapted from Miyo *et al*., 2017.

Additionally, due to the prior assignment of SLC25A32 as the mitochondrial folate transporter, enzymes involved in folate metabolism were analysed (Appendix 7.6). Folate metabolism is compartmentalised into the nucleus, cytoplasm and mitochondria (Nijhout *et al*., 2006), Figure 4.19. Of the mitochondrial enzymes detected the bifunctional methylenetetrahydrofolate dehydrogenase MTHFD2 showed no appreciable change; serine hydroxymethyltransferase (SHMT2) and monofunctional C1-tetrahydrofolate synthase (MTHFD1L) also showed no change, although there was discrepancy between the labelling orientations. Therefore, from this data, there is little indication that SLC25A32 knockout caused changes in mitochondrial folate-related enzyme abundance.

4.7. SLC25A32 knockout causes cell wide transcriptional changes

Total RNA was purified from conditioned HAP1 and SLC25A32 knockout cells as biological triplicates and analysed by transcriptome profiling (Eurofins NGS Favourite Transcriptomics Service). The transcripts detected, measured as transcripts per million (TPM), were processed by Dr Alan Robinson using DESeq2 (Love *et al*., 2014) to obtain the fold change and adjusted p-value for the expression of each gene in the SLC25A32 KO compared to the HAP1 control. Adjusted p-values take into account both the level of change in expression and the variability between multi-sample testing, therefore this value was used to sort the data for interpretation, rather than using the fold change alone, with a cut-off of P <0.05 to denote significant differential expression.

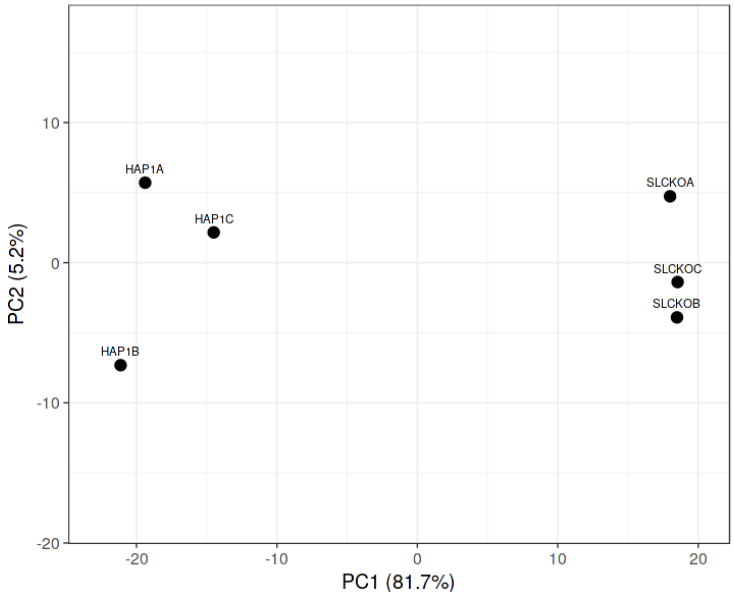


Figure 4.20 Principal component analysis (PCA) plot. The TPM for the 500 most differentially expressed genes were assessed for each sample using the online platform ClustVis (Metsalu and Vilo, 2015). For the six samples the three knockouts were more similar to one another than to the controls. The first component, PC1, accounted for 81.7% of the variance and overall both components represented 86.9%, demonstrating the reliability of the triplicates.

Figure 4.20 shows a PCA plot for the data, with most variance between the sample types than between the triplicates. Heat mapping of differentially expressed genes in Figure 4.21B also demonstrates the grouping of the triplicates and the difference in expression profiles between the cell lines. Transcriptomics resulted in the detection of 11965 gene transcripts, of which 2054 were significantly differentially expressed (P <0.05) - 51% of these showed decreased expression in the knockout. Interestingly only 8% of these were assigned mitochondrial proteins, which represented 16% of the total mitochondrial protein gene transcripts detected. When compared to the proportion of non-mitochondrial proteins that were differentially expressed (17%) this result implied there was no difference between the level of expression change in mitochondrial protein genes compared to other cellular proteins. The transcriptome data was assessed for changes in expression of proteins of interest, particularly those involved in oxidative phosphorylation, riboflavin handling and folate metabolism. As expected a large

number of transcription factors were among the most differentially expressed genes. Regulating multiple pathways could be a strategy to implement large-scale changes in gene expression – 22% of the top 50 most differentially expressed were transcription factors (raw data in Appendix 7.7).

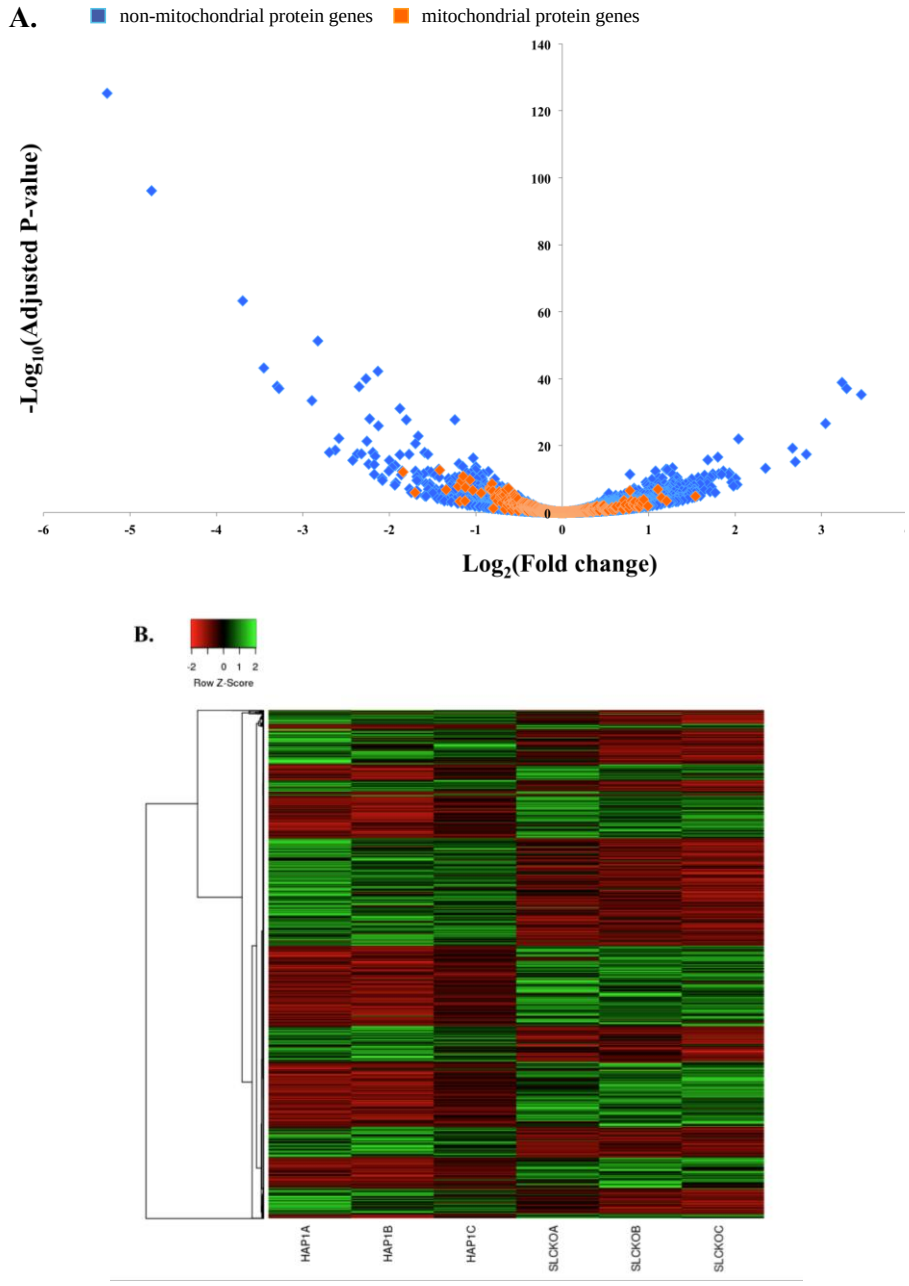


Figure 4.21 Transcriptional differences between HAP1 and SLC25A32 KO cell lines. A. Volcano plot of all detected genes. Data are plotted on as Log_2 for fold change and as $-Log_{10}$ for adjusted P-value. Adjusted P-value was used to determine significantly differentially expressed genes, P <0.05, with differentially expressed mitochondrial protein genes shown in orange and non-mitochondrial protein genes shown in blue. Non-significantly expressed genes from the two groups are shown in light orange and light blue respectively. Mitochondrial proteins were assigned using the IMPI database (version Q3, 2017; Smith & Robinson, 2016). **B. Heatmap showing the expression profile of each sample for all genes with adjusted P-value <0.05.** Data was plotted using TPM for each sample using the online platform Heatmapper (Babicki *et al*., 2016). Green bars show up-regulated genes and red down-regulated genes - clustering demonstrates that the HAP1 and SLCKO triplicates have a distinct profile of differentially regulated genes.

4.7.1. Transcriptional changes to SLC25A32

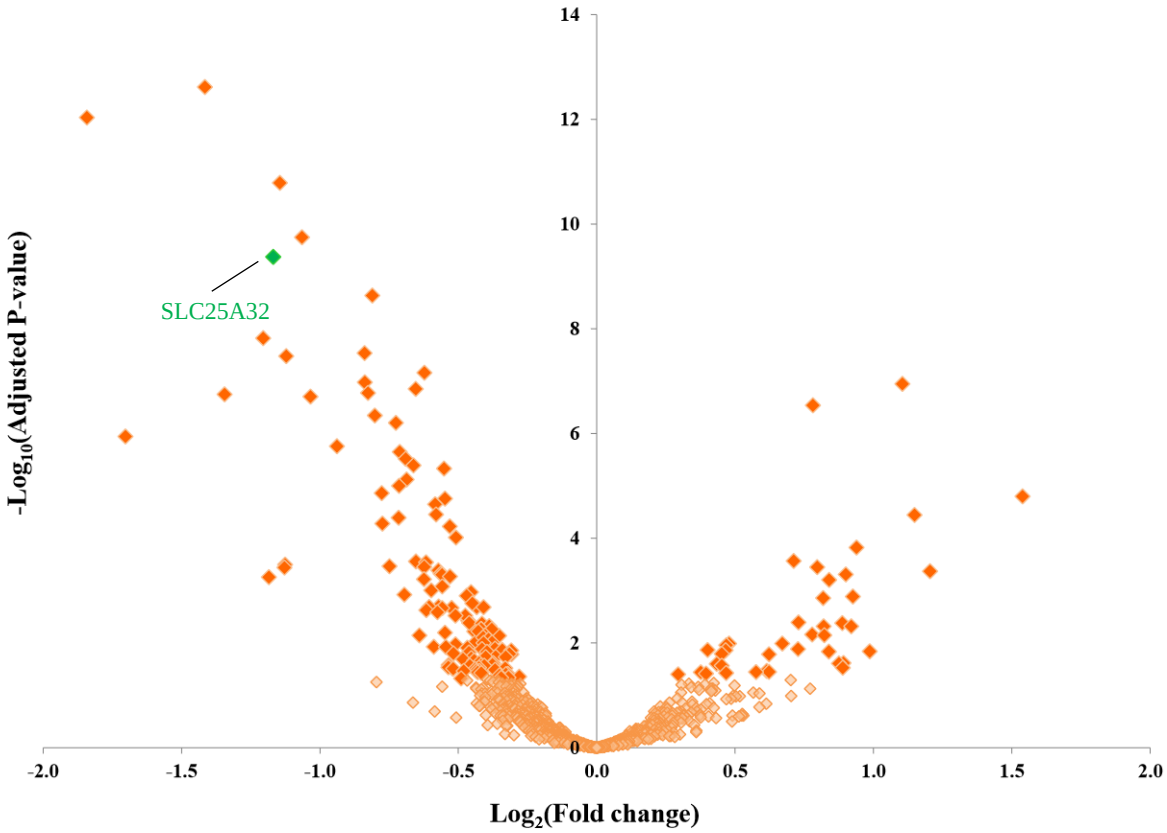


Figure 4.22 Volcano plot of all genes encoding mitochondrial proteins, showing SLC25A32 was one of the most differentially expressed of these genes. Data are plotted on as Log_2 for fold change and as $-Log_{10}$ for adjusted P-value. Adjusted P-value was used to determine significantly differentially expressed genes, $P < 0.05$, with differentially expressed mitochondrial protein genes shown in orange and non-significantly expressed mitochondrial protein genes in light orange. SLC25A32 is highlighted in green Mitochondrial proteins were assigned using the IMPI database (version Q3, 2017; Smith & Robinson, 2016).

As shown in Figure 4.22, SLC25A32 was detected as one of the most differentially expressed mitochondrial protein encoding genes; in the knockout cell line mRNA were detected at 2.25-fold lower levels. To investigate this further the coverage across the whole gene was assessed and analysed by exon, with read depth for the first three exons shown in Figure 4.23. CRISPR resulted in a frameshift deletion in exon 2 that introduced a premature termination codon (PTC, section 4.4.2), and due to its location in an early exon is known to trigger nonsense-mediated decay (Popp & Maquat, 2016). Therefore only low steady-state levels of abberant transcripts are likely, due to their rapid degradation, yet the transcriptome demonstrated coverage across the whole of *SLC25A32*. However read depth was greatly decreased, in the HAP1 control reads/base of up to 300 were detected but in the knockout values did not exceed 100, Figure 4.23. Interestingly there was a similar magnitude of decrease in reads (around three-fold) across theentire gene – exemplified by exons 1 and 3 in Figure 4.23 – and so the overall trend in covergae was retained. Reads across exon 1 were comparatively low in both cell lines, particularly at the 5' end, with higher reads across exon 3 in both cell lines. In contrast, in HAP1 the coverage across exon 2 was similar to that of exon 3, whereas SLCKO showed a disproportional decrease with a

complete absence of coverage across the region targeted by CRISPR, Figure 4.23. Although this was unexpected, given that the gene should have been knocked out, there have been reports of unintended outcomes when using CRISPR to affect genomic level knockout. Xue and co-workers demonstrated that sgRNA causing a frameshift indel resulted in a moderate decrease in mRNA levels of the oncogene *Kras* and specifically fewer reads across exon 2, which contained the deletion, in adrenocarcinoma mouse cells. Exon-skipping was identified as responsible due to the presence of alternative splice variants, confirmed by reverse transcription PCR (Mou *et al*., 2017). This has also been demonstrated in HeLa cells, a single base insertion in *FLOT1* resulted in mRNA corresponding to multiple splice variants, indicating that the insertion caused alternative splicing, therefore transcripts included different exons, although the correct exon-intron boundaries were used (Kapahnke *et al*., 2016). Consequently, in the same way that SLC25A32 peptides were detected by SILAC in the knockout cell line, transcriptomics also suggests that the CRISPR-mediated deletion did not confer a complete SLC25A32 knockout.

Figure 4.23 Transcript analysis for HAP1 and SLCKO across the first three exons of SLC25A32 clearly shows the deleted region in exon 2. Reads per base for HAP1 are shown in blue and SLCKO in pink aligned with the cDNA sequence. HAP1 shows good coverage with read depth up to 300, whereas SLCKO shows overall lower coverage across all three exons (<100) with the greatest decrease observed for exon 2. The gap in the sequence corresponds to the deleted region in the CRISPR knockout.

4.7.2. Folate metabolism and riboflavin-handling pathways

The most differentially expressed gene in the knockout was ALDH1A2, an aldehyde dehydrogenase, which was down-regulated by over 30-fold. This hormone signalling protein catalyses the synthesis of retinoic acid and is key to development, but is not a flavoprotein and has no direct role in either folate or riboflavin metabolism. However there may be a link between retinoic acid and increased folate receptor expression, as investigated in cultured mouse and human cells (Bolton *et al*., 1999). The folate receptor is responsible for folate uptake by the cell and therefore changes to its expression would profoundly affect cellular folate metabolism. Folate receptor transcripts were detected in the transcriptome analysis, however no significant change was observed – therefore the significance of altered ALDH1A2 expression in the knockout remains unclear. However it was striking that the second most differentially expressed gene is directly involved in folate metabolism. The plasma membrane bound folate hydrolase, FOLH1, was down-regulated by over 16-fold in the knockout. Despite being upstream in the folate-handling pathway from SLC25A32, its lowered expression suggests that folate movement within the cell is altered in the knockout. However FOLH1 is also involved in alanine, aspartate and glutamate metabolism, which is more closely linked with mitochondria and the citric acid cycle - therefore this change in expression could be related to wider mitochondrial changes. The transcripts of nine other folate metabolising enzymes were detected – four cytosolic (DHFR, MTHFR, MTHFD1 and SHMT1) protein genes displayed no change in expression, along with MTHFS, MTHFD1L, MTHFD2 and SHMT2 that are localised to the mitochondria. However mitochondrial MTHFD2L showed differential expression, although the increase in expression was low (less than two-fold). Thus on a transcriptional level there was no consistent change in levels of folate metabolising enzymes as a result of SLC25A32 knockout.

Similarly proteins involved in riboflavin handling were assessed for transcriptional regulation. The plasma membrane transporter SLC52A2 and RFK were both unaffected by the knockdown, however the FADS gene (FLAD1), which encodes both a cytoplasmic and mitochondrial isoform was significantly reduced ($P = 1.778 \times 10^{-10}$). FLAD1 was down-regulated by more than two-fold and as the FADS1 isoform is present downstream of mitochondrial transport this may represent some regulation by the cell, to decrease FAD production in response to lowered mitochondrial flavin concentration; although both isoforms would be affected.

4.7.3. Mitoflavoproteins

The transcription profiles of 28/32 known human mitoflavoproteins were detected, however only two, dimethylglycine dehydrogenase (DMGDH) and NDUFV1, showed a significant difference in expression. Unusually DMGDH expression was increased, which is counter-intuitive since this enzyme requires an FAD cofactor. However it is involved in amine metabolism and therefore its transcriptional

regulation may be dominated by changes in this pathway. Once again the transcriptome data indicates that any changes to the mitoflavoproteome in SLC25A32 knockout cells, as suggested by the SILAC data (Figure 4.18), are not due to gene regulation - protein abundance may be affected by decreased stability or increased turnover.

4.7.4. Transcriptional changes in complex I genes

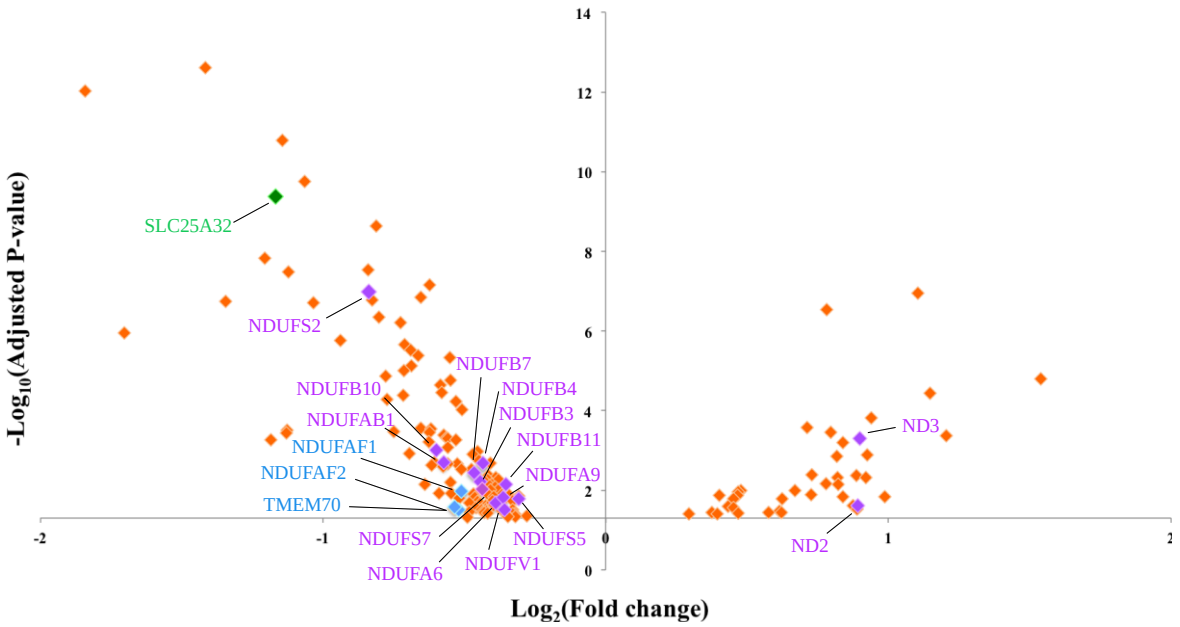


Figure 4.24 Volcano plot for differentially expressed mitochondrial protein genes. Data are plotted as Log_2 for fold change and as $-Log_{10}$ for adjusted P-value. Adjusted P-value was used to determine significantly differentially expressed genes, $P < 0.05$, and the y-axis cut-off set at this value. Complex I subunit genes are shown in purple, assembly factor genes in blue and SLC25A32 in green (acting as a positive control).

Transcripts for all 44 complex I subunit genes were detected but only 32% were differentially expressed; 12 were down-regulated and two up-regulated (Figure 4.24). ND2 and ND3 showed significantly increased expression, whilst all other mtDNA-encoded subunits also displayed transcript increases in the knockout. This contrasts to the nuclear-encoded subunits, which, apart from NDUFS4 and NDUFC2, all demonstrated decreased expression. It is difficult to interpret this observation, especially given the lack of SILAC data for the ND subunits, but it may indicate some difference in regulation of the mitochondrial-encoded respiratory chain subunits compared to the nuclear-encoded subunits in the knockout; especially given the differential expression of the two mtDNA-encoded ATP synthase genes (Table 4.1). However the transcripts from the mtDNA-encoded complex III and IV genes were not detected.

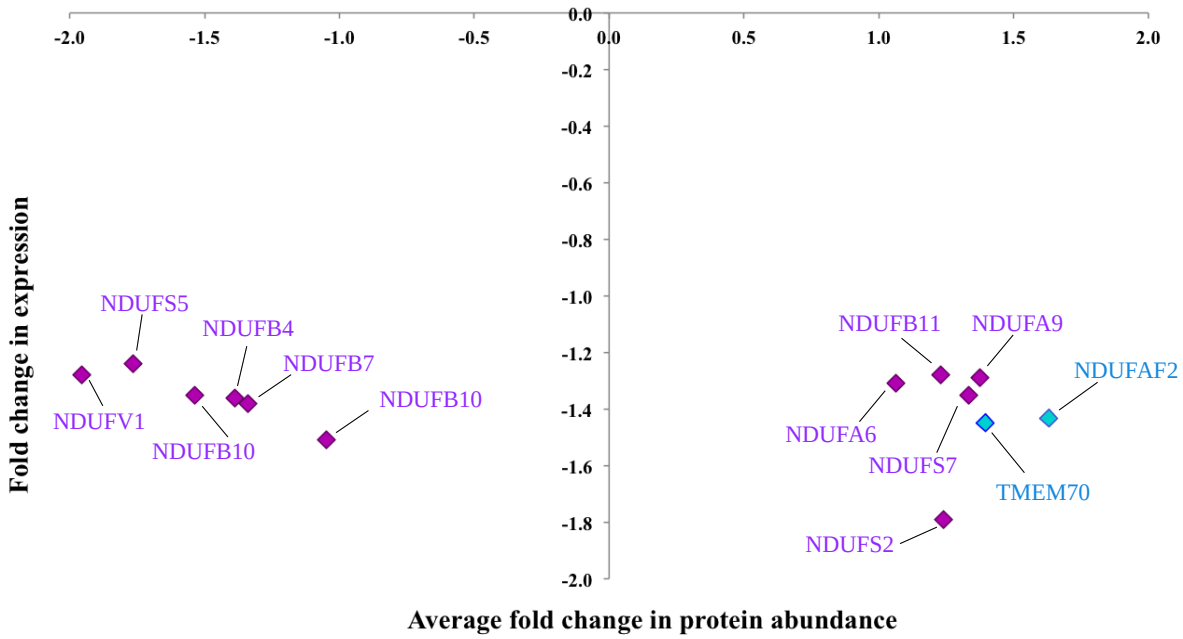


Figure 4.25 Expression and abundance comparison for differentially expressed complex I subunit and assembly factor genes. 12 subunits and three assembly factors were significantly differentially expressed in the transcriptomics data (P <0.05) – 11 subunits (purple) and two assembly factors (blue) were detected by SILAC. Negative values denote decreased expression or abundance in the SLC25A32 KO compared to the HAP1 control. To compare the two data sets the forward and reverse values for fold change in abundance from the SILAC (section 4.6.2.2) were averaged and plotted against fold change in expression. This demonstrated that subunits separated into two cohorts, those with correlating expression and abundance and those with no consensus.

Although 12 complex I subunits displayed a significant decrease in expression none of these changes resulted in a large fold decrease and when compared to their structural location there was little correlation. Despite NDUFV1 showing significant down-regulation its neighbouring subunits (NDUFS1, NDUFV2, NDUFA2) did not display considerable changes. To further investigate whether complex I subunits were under altered transcriptional control, the transcriptome (gene expression) and SILAC (protein abundance) data from the differentially expressed subunits were compared. Figure 4.25 shows that the 11 subunits detected in both experiments separated into two groups – six showed both lowered expression and abundance whilst five only showed decreased expression. Changes in protein abundance of NDUFV1, NDUFS5 and NDUFB4 (shown by a greater than 1.5-fold decrease in SILAC) are likely to result from down-regulation of gene expression, as transcripts detected were also decreased by greater than 1.5-fold in the knockout. However, in general, there was little consensus between the transcription profile and protein abundance of complex I subunits, which would indicate that changes in protein levels observed by SILAC were independent of transcriptional control. Additionally, the three complex I assembly factors with significant down-regulation (Figure 4.24: NDUFAF1, NDUFAF2 and TMEM70) also demonstrated little correlation with the SILAC data (NDUFAF2 and TMEM70 in Figure 4.25). These assembly factors are involved in building different subcomplexes, making it difficult to rationalise the changes in expression observed.

4.7.5. Differential expression of respiratory chain genes

Gene	Chromosome	Fold change	Adjusted P-value
NDUFS2	1	1.79	1.05×10^{-7}
MT-ND3	MT	1.87	4.92×10^{-4}
NDUFB10	16	1.51	9.85×10^{-4}
NDUFAB1	16	1.49	2.01×10^{-3}
NDUFB4	3	1.35	2.07×10^{-3}
NDUFB7	19	1.38	3.70×10^{-3}
NDUFB3	2	1.36	5.99×10^{-3}
NDUFB11	X	1.28	7.27×10^{-3}
NDUFS7	19	1.35	9.56×10^{-3}
NDUFA9	12	1.29	1.52×10^{-2}
NDUFS5	1	1.24	1.65×10^{-2}
NDUFA6	22	1.31	2.10×10^{-2}
MT-ND2	MT	1.86	2.41×10^{-2}
NDUFV1	11	1.28	3.10×10^{-2}
SDHC	1	1.58	4.07×10^{-6}
UQCR11	19	1.39	1.26×10^{-3}
UQCRC1	3	1.31	4.75×10^{-3}
UQCR10	22	1.34	5.33×10^{-3}
UQCRH	1	1.32	1.48×10^{-2}
COX5A	15	1.54	6.93×10^{-8}
COX6B1	19	1.30	5.46×10^{-3}
COX5B	2	1.38	1.52×10^{-2}
COX4I1	16	1.24	1.62×10^{-2}
COX6A1	12	1.26	1.86×10^{-2}
ATP5G1	17	1.46	1.74×10^{-5}
ATP5J2	7	1.44	5.92×10^{-5}
ATP5I	4	1.48	4.73×10^{-4}
ATP5J	21	1.33	2.07×10^{-3}
MT-ATP6	MT	1.77	4.78×10^{-3}
ATP5S	14	1.39	1.02×10^{-2}
MT-ATP8	MT	1.66	1.30×10^{-2}
ATP5G3	2	1.34	2.65×10^{-2}
ATP5H	17	1.21	4.39×10^{-2}

Table 4.1 Differentially expressed respiratory chain subunit genes. Subunits from each of the five complexes that were differentially expressed between HAP1 and SLC25A32 KO lines are shown. Genes are ordered within each complex by adjusted P-value. Fold change is highlighted in red for down-regulation and green for up-regulation in the SLC25A32 knockout.

Although some subunit genes from each of the complexes were significantly differentially expressed, as shown in Table 4.1, the change in expression was below two-fold, consistent with the lack of large changes in protein abundance, determined by SILAC. However, the expression of two overlapping mtDNA-encoded subunits of complex V, ATP6 and ATP8, were significantly up-regulated, similarly to mtDNA-encoded ND2 and ND3 of complex I. ATP8 displayed an increase in expression but considerable decrease in protein abundance, although (as previously stated) it is a small hydrophobic protein and likely to be inconsistently detected by mass spectrometry. Comparing transcriptome and SILAC data (Figure 4.13) suggests that the differences in complex II subunit levels (SDHA and SDHB)

arose at protein level, from decreased stability or increased turnover, rather than from expression regulation, as there was no change in expression, although SDHC showed some down-regulation. This indicates that the lack of flavin has a direct effect on the SDHA protein and perhaps, due to its structural association, SDHB is also detrimentally affected.

Importantly, complex III subunits displayed a noticeable shift towards lowered abundance in SLC25A32 knockout cells, as shown by Figure 4.17. This is particularly interesting because the complex does not contain a flavin cofactor, therefore depleted flavin would not be expected to directly affect the protein. However, complex III subunits also demonstrated lowered transcript levels; Table 4.1 shows that UQCRC1, UQCR10, UQCR11 and UQCRH were differentially expressed and transcripts for all eight complex III subunits detected were down-regulated to some extent. Transcriptional regulation may therefore result in lowered complex III expression, which is also detected as decreased protein abundance. This could reflect a cellular response to flavin depletion and disruption of complexes I and II. Thus, the reason for changes in complex III abundance differs from the impact on complex II, which is likely to be caused by direct protein disruption due to the lack of flavin, and suggest that changes to complex III were instigated as a cellular feedback response, on a genomic level.

4.7.6. The respiratory chain was the most down regulated pathway in the SLC25A32 knockout cells

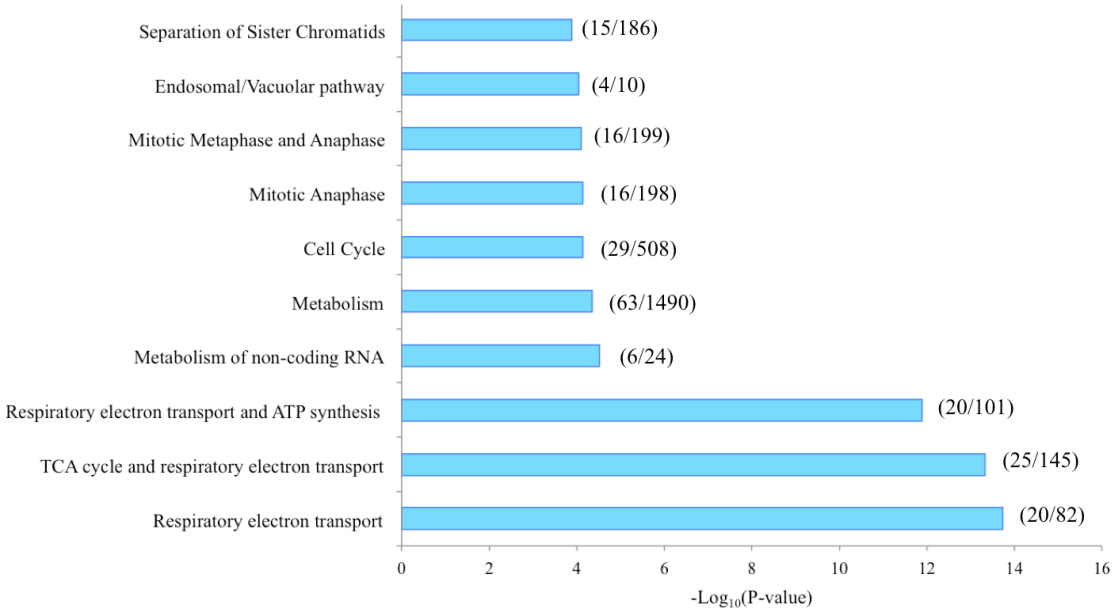


Figure 4.26 Protein networks down regulated in SLC25A32 knockout cells. Differentially expressed genes (adjusted P-value <0.05) were assessed by protein interactions to explore changes in cellular pathways. Analysis was done using the online platform Network Analyst (Xia *et al*., 2015). The top ten most down regulated pathways by P-value are shown using the STRING interactome with a zero order network and the Reactome database. Numbers in parenthesis denote the number of genes down regulated compared to the total number in the network.
Although it is informative to investigate changes in expression for certain gene sets, particularly those associated with the putative function of SLC25A32 and complex I, it is also interesting to adopt a non-

biased approach and explore changes in protein networks as a result of SLC25A32 knockout. This was achieved using Network Analyst (Xia *et al*., 2015) to assess the differentially expressed genes and arrange them into an interaction network, using the STRING interactome. Figure 4.26 demonstrates the ten most down-regulated pathways, clearly showing the respiratory chain as the most suppressed pathway across the SLCKO cells. The top three networks incorporate slightly different proteins but all centre on the four respiratory chain complexes. Results using the Reactome protein network also highlights general metabolism as being down-regulated, which although non-specific does imply that large-scale changes have been instigated by the knockout. There has also been considerable disruption to the cell cycle and the endosomal pathway of the major histocompatibility complex.

4.8. Complexomics demonstrated that the N-domain of complex I was still assembled in the mitochondria of SLC25A32 knockout cells

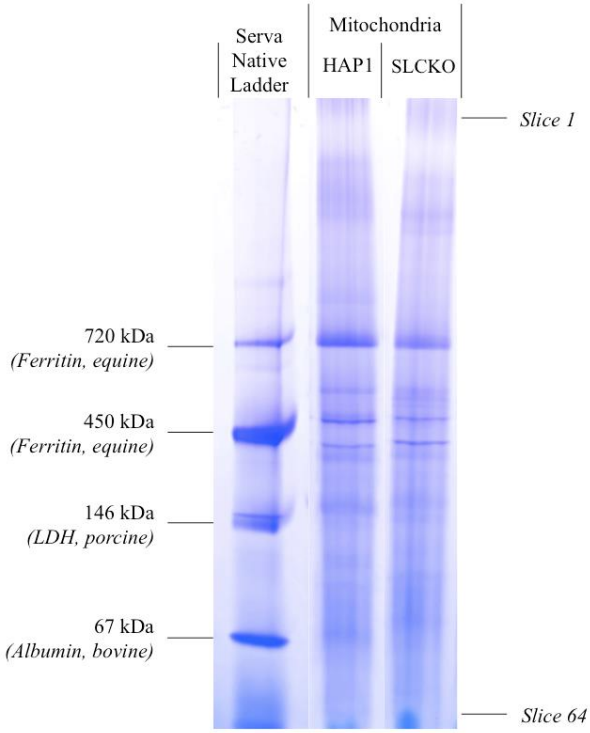


Figure 4.27 Mitochondrial samples from conditioned HAP1 and SLCKO cells used for complexome analysis. Mitochondria were purified from cells lines, as described in Chapter 2, section 2.3.3, and solubilised in 1.2% DDM prior to BN-PAGE. 10 µg protein was loaded and the Serva Native ladder used to indicate protein masses. Once run the gel was stained with Coomassie before submission for complexome analysis - both the sample lanes were cut into 64 even slices - direction of slice numbering shown.

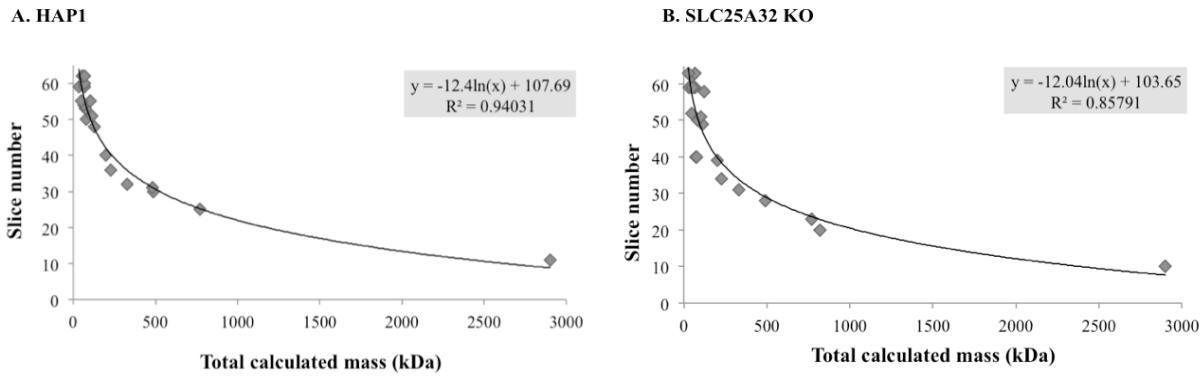


Figure 4.28 Protein mass calibration curves for mitochondrial complexomes. Although both samples were run on the same gel for better comparison of complexes and their estimated masses, calibration curves were constructed. Proteins and protein complexes of known masses were chosen to determine molecular weight migration values in each gel slice. **A. HAP1. B. SLCKO.** Protein masses were obtained from the UniProt database (UniProt Consortium T, 2018). Table of proteins used in Appendix 7.8A&B.

Complexomics was used to build on SILAC and transcriptomics data by investigating the structure of complex I, and other respiratory complexes, in the SLC25A32 knockout (raw data in Appendix 7.8C&D). Figure 4.27 shows the BN-PAGE of 10 µg mitochondria samples for SLCKO and the control HAP1. The calibration curves in Figure 4.28 were used to determine the estimated mass of complexes found in each slice (table of proteins used in Appendix 7.8A&B). Slice 44 of the HAP1 complexome failed during mass spectrometry measurements and so was omitted from analysis; slice 64 from the SLCKO complexome was also unusable – so the last slice of both complexomes was removed.

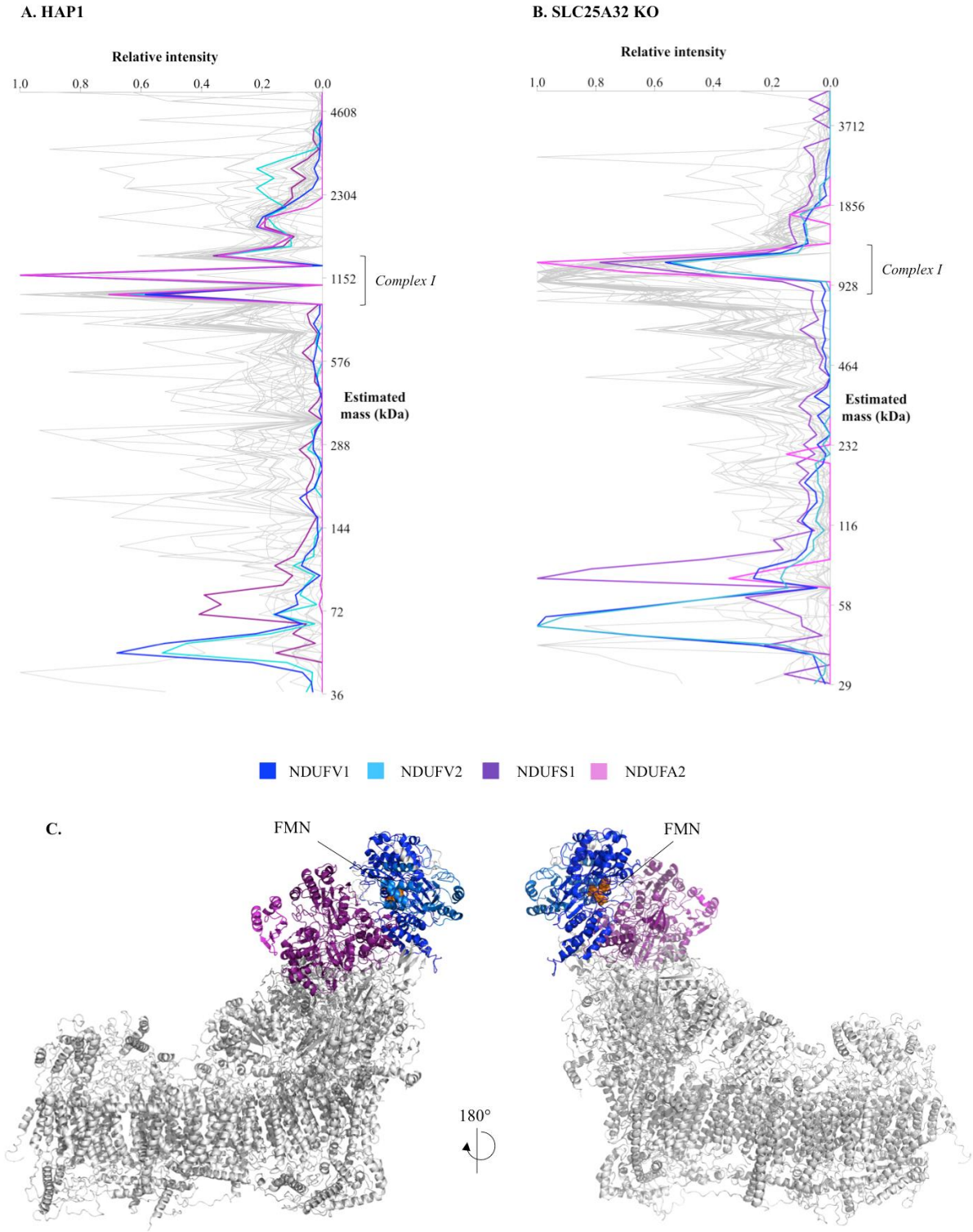


Figure 4.29 Distribution of complex I across gel slices after BN-PAGE. A. HAP1 and **B. SLCKO** migration profiles of complex I subunits. Distinct populations of complex I subunits were observed in the HAP1 control, whereas in the knockout the position of mature complex I was spread over multiple slices as a broad peak. The four subunits of the distal N-domain are highlighted: NDUFV1 in navy, NDUFV2 in blue, NDUFS1 in purple and NDUFA2 in pink. **C.** Position of the four subunits in the distal N-domain, coloured to match the profiles. The position of FMN is shown in orange. The human model of complex I was constructed using mouse complex I cryo-EM structures (Chapter 6, section 6.6).

Figure 4.29A and B show the clustering profiles of complex I subunits for HAP1 and SLCKO, respectively. 42/45 subunits were detected in HAP1 mitochondria (missing ND3, ND4L and NDUFB2) and 41/45 subunits were detected in SLC25A32 knockout mitochondria (missing ND3, ND4L, NDUFB2 and NDUFV3). Unlike previous analysis of complexomes from 143B cells (Chapter 3, section 3.8.1, Figure 3.22) the complex I profile in HAP1 cells did not show a single complex I population; instead the HAP1 control showed a dual peak profile ~1000 kDa. This has not been observed previously, and although the profile may represent two populations of complex I, no differences could be found to account for the migration difference – the same number of subunits and low levels of assembly factors were present in each. The experiment was repeated and the same profile obtained for complex I in the HAP1 control, therefore the cause of this unusual distribution is yet to be determined. The SLC25A32 knockout cell line, in contrast, displayed a more typical complex I profile of a single broad peak spanning several slices at 958-1131 kDa. Despite some uncertainty over the complex I distribution from the control complexome, information on the flavin containing N-domain could still be ascertained. As shown in Figure 4.29, the migration of the four subunits (NDUFV1, NDUFV2, NDUFS1 and NDUFA2) comprising the distal region of the flavin (N-) domain was analysed. These subunits are the final subcomplex to be assembled onto complex I during biogenesis. As previously discussed, SLCKO cells showed no observable complex I activity by in-gel activity (section 4.5.3.1) and compromised OXPHOS (section 4.5.1), which raised the question of whether the N-domain is assembled in the absence of flavin. Lack of complex I activity could arise from the enzyme missing the FMN containing N-domain, or perhaps just lacking the cofactor. The distribution of these subunits provides insight into when FMN is inserted into NDUFV1 and the impact of this on the stability of the enzyme. Complexome analyses indicate that despite lack of FMN the N-domain is correctly assembled, with NDUFV1, NDUFV2, NDUFS1 and NDUFA2 present in slices corresponding to complex I in the SLCKO profile, Figure 4.29B. Compared to the HAP1 control, Figure 4.29A, there is little difference in the migration patterns of these subunits, they are both present at around 1000 kDa, as part of the complete enzyme, and at low molecular weights in their monomeric forms. Although there appears to be slightly lower relative intensities of these subunits associated with complex I in the knockout, and therefore more in the free form, overall, even in the absence of the flavin cofactor, all complex I subunits are assembled. The difference between complex I in HAP1 and SLCKO cells, accounting for the difference in activity, is therefore likely to be only the missing cofactor, rather than a stalled, incompletely assembled enzyme. This is new information and suggests that FMN insertion into NDUFV1 may be the final stage of assembly, taking place once all protein subunits are present and may correlate with the removal of assembly factors to release active complex I.

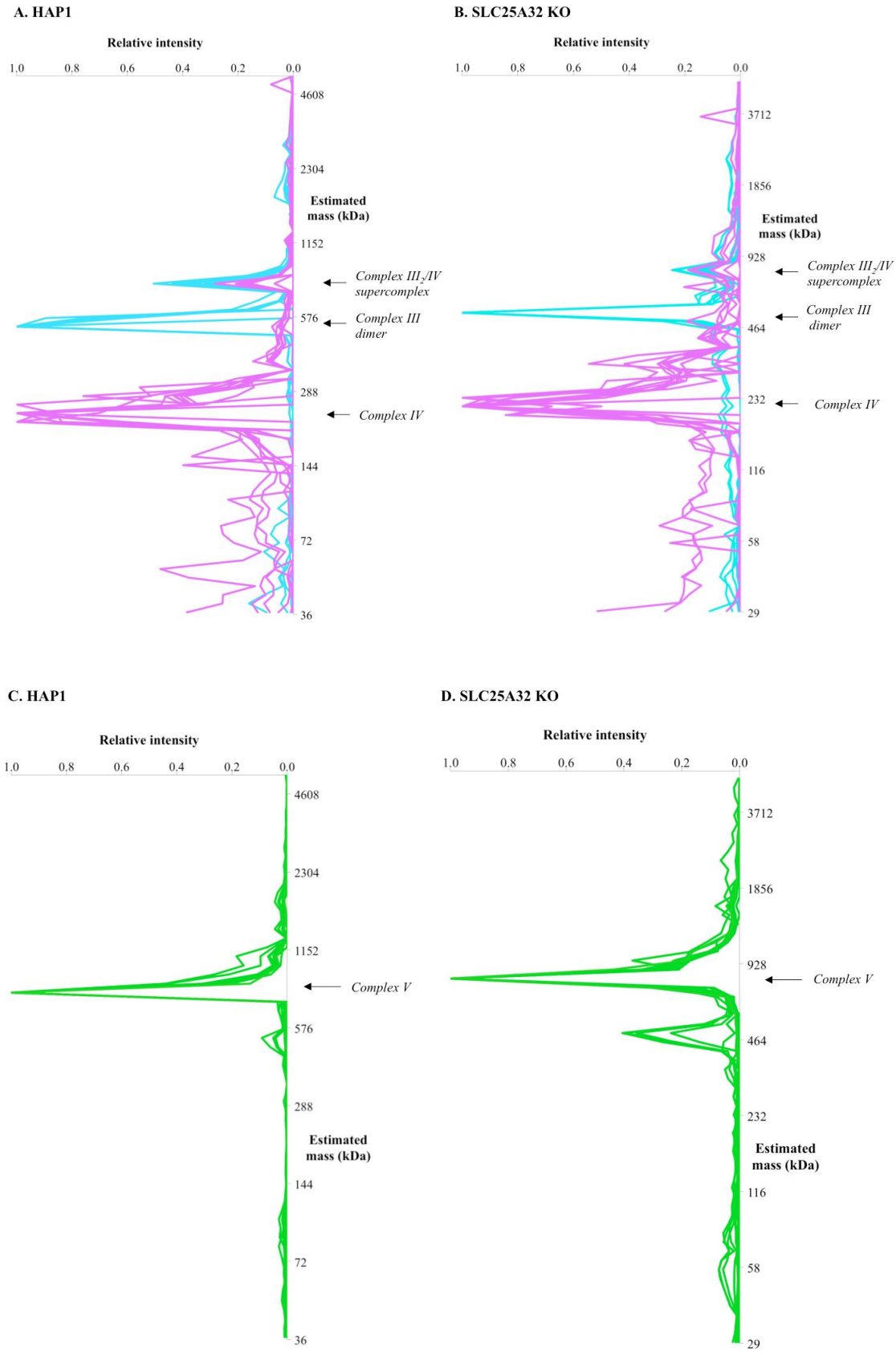


Figure 4.30 Distribution of respiratory complexes III, IV and V across gel slices after BN-PAGE. A. HAP1 and **B. SLCKO migration profiles of subunits of complexes III and IV.** Complex III subunits are shown in cyan and complex IV subunits shown in magenta. **C. HAP1** and **D. SLCKO migration profiles of subunits of complex V.** Complex V subunits are shown in green. Profiles for different complexes are shown in separate panes of clarity.

The migration profiles of subunits from complex III and IV are shown in Figure 4.30A and B; both complexes displayed a similar profile in HAP1 and SLCKO. The complex III dimer migrated to 526 kDa (calculated mass 490 kDa) in HAP1 with nine subunits detected and all subunits were also found to migrate to 787 kDa with lower intensities, indicating a subpopulation in the complex III_2/IV supercomplex. Complex III appeared unaffected in SLCKO mitochondria; eight subunits were detected with an estimated mass of 536 kDa for the dimer. The complex III profile clearly indicates the presence of an intact population of complex III, with one high-intensity peak, and another larger molecular mass peak of lower intensity. This peak, at 811 kDa in SLCKO mitochondria, represents the III_2/IV supercomplex (calculated mass 690 kDa) and although found at a larger mass than expected it contains subunits from both complexes III and IV, which do not co-migrate elsewhere. Lower relative intensities of complex III subunits were found in the supercomplex from SLCKO, compared to the control. It is unclear why SLC25A32 knockout would detrimentally impact complex III association with complex IV, however SILAC showed complex III subunits were decreased in abundance, Figure 4.17 – and therefore may reflect a change in proportions of complex III assemblies due to decrease in overall protein levels. 12 complex IV subunits were detected in the HAP1 samples and 11 in SLCKO, with both showing a peak in relative subunit intensities around the expected mass of the intact complex (200 kDa) – 235 kDa and 215 kDa respectively – although the highest intensity of all subunits was not found in a single slice. Unusually the SLCKO profile also highlighted an additional low intensity peak in complex IV subunits at 325 kDa, this represents a slight difference compared to the control. However, the presence of complex IV subunits within a supercomplex with complex III, at similar relative intensities in both samples suggests that in general this enzyme has not been substantially affected by the knockout. This supports the in-gel activity shown in Figure 4.12.

The distribution of ATP synthase subunits was also assessed; they showed similar migration profiles for HAP1 and SLCKO, with a single peak containing the same 14 subunits at 787 kDa and 811 kDa respectively (calculated mass 770 kDa), Figure 4.30C and D. Additionally both complexomes showed a smaller molecular weight peak, (526 kDa in HAP1 and 493 kDa in SLCKO) containing moderate relative intensities of α, β, γ and δ subunits, which are all part of the F_1 domain and could indicate the presence of subassemblies (Walker *et al*., 1985). This peak contained greater relative intensities of these subunits in the knockout, which may indicate a higher abundance of complex V assembly intermediates compared to the control. However, in both cell lines mature complex V was present and all subunits showed highest intensity in the intact complex.

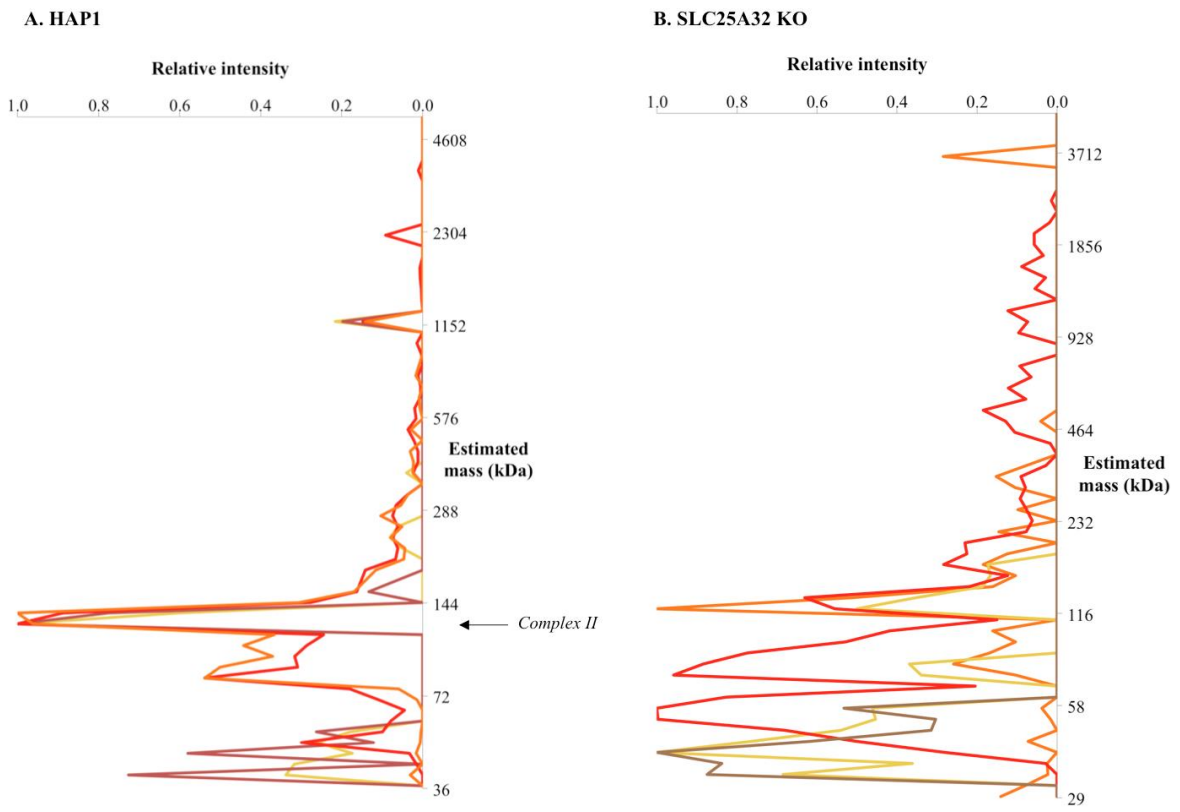


Figure 4.31 Distribution of complex II across gel slices after BN-PAGE. A. HAP1 and **B. SLCKO migration profiles of subunits of complex II.** Subunits are shown in different colours: SDHA in red, SDHB in orange, SDHC in yellow and SDHD in brown.

In contrast, the profile of complex II differed considerably between HAP1 and SLCKO, Figure 4.31. In the control there was a peak in all four subunits at 123 kDa, indicating complex II (calculated mass 130 kDa) and a smaller peak at high molecular mass (1178 kDa), which may represent complex II in a supercomplex. However in the knockout the four subunits did not comigrate; a peak in SDHB relative intensity at 120 kDa showed only moderate relative intensities of the FAD containing subunit SDHA and SDHC, whereas the last subunit, SDHD, was absent and only found at low molecular weights, indicating a monomeric form. This suggests while HAP1 contained active succinate dehydrogenase, along with the subunits in their free forms, the SLC25A32 knockout did not contain mature complex II. The highest relative intensity of SDHA in the knockout was at 57 kDa and likely to correspond to monomeric protein (calculated mass of 72 kDa). Therefore a complete loss of complex II and the presence of subunits as monomers support the lack of complex II activity observed in section 4.5.3.2.

4.9. CRISPR deletions in exon 1 of SLC25A32 to further validate the effect of SLC25A32 knockout

CRISPR was used to generate knockout HAP1s with deletions in exon 1 as discussed in Chapter 2, section 2.3.3. These were designed to provide additional confirmation of the role of SLC25A32 in parallel with the commercially obtained SLC25A32 KO (SLCKO).

Figure 4.32 Sequencing data showing successful deletion in SLC25A32 exon by CRISPR. Two colonies treated with sgRNA-1 construct were found to contain a deletion in the correct region. Sequence alignment of SLC25A32 (top line) and knockout cell lines (below) show exon 1 deletions. Highlighted in orange is the deleted region, CRISPR KO-1 demonstrated a 26 bp deletion and CRISPR KO-2 a 7 bp deletion. Both resulted in a frame shift that also brought a stop codon into frame, highlighted in red. The location of the sgRNA-1 target is underlined in green.

FACS positive colonies were cultured and screened for deletions by PCR (Chapter 2, section 2.3.3). Two colonies were identified from treatment with the sgRNA-1 vector, deletions are shown in Figure 4.32. These cells were then conditioned and incucyte data collected to investigate cell proliferation and morphology, with respiratory profiles obtained by extracellular flux analysis.

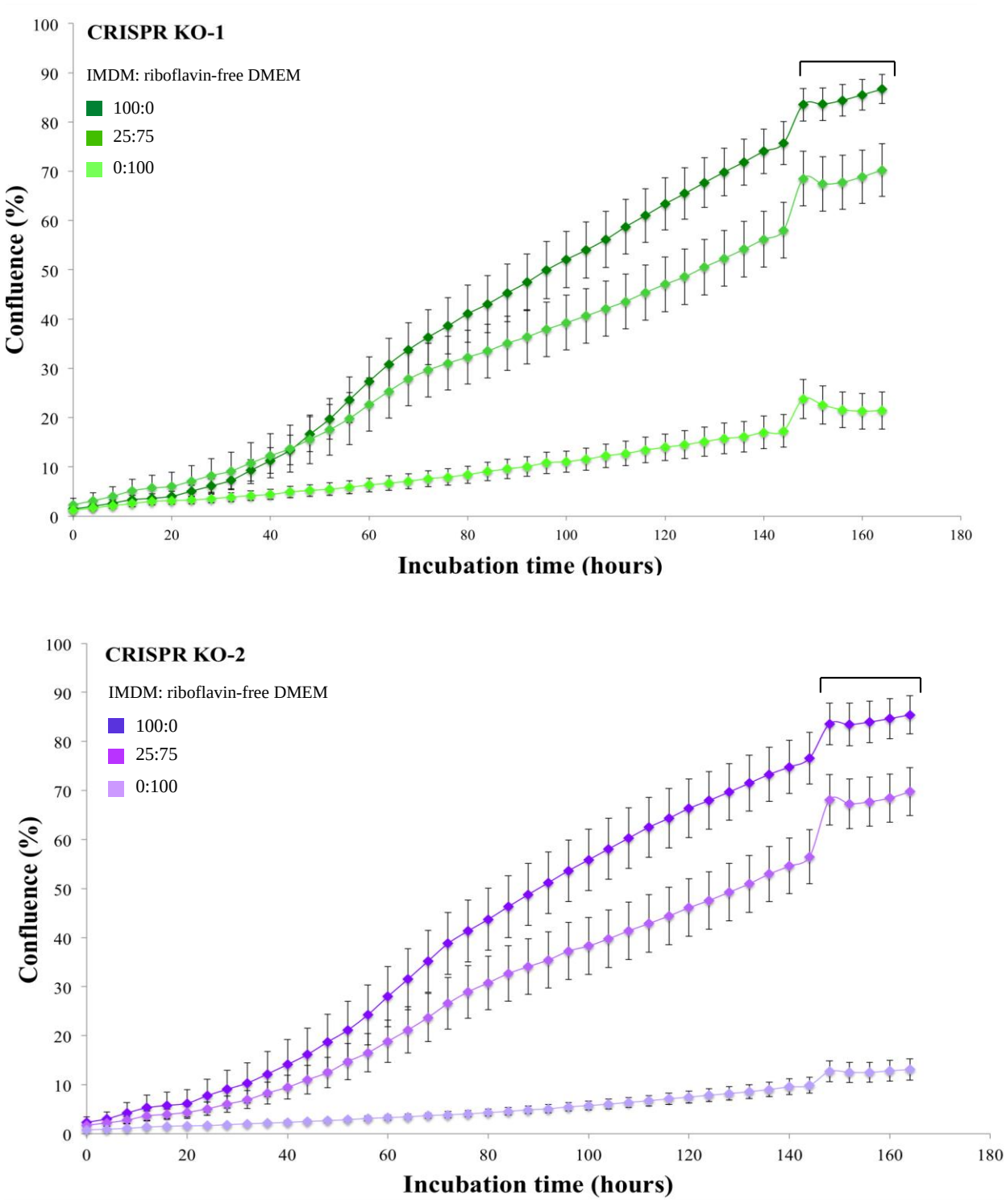


Figure 4.33 Growth profiles of CRISPR knockouts during conditioning stages. Cells were seeded at 14,000/well with 4 ml of appropriate medium and incubated at 37 ºC for at least 160 hours. Incucyte software was used to record the change in growth, with images obtained every 4 hours and percentage confluence calculated. Error bars denote standard error. Cells were adapted from IMDM to riboflavin free DMEM in 25% increments after two passages. CRISPR KO-1 shows IMDM grown cells in dark green, IMDM: riboflavin free DMEM 25:75 in lighter green and riboflavin-free DMEM in bright green. CRISPR KO-2 shows IMDM grown cells in dark purple, IMDM: riboflavin free DMEM 25:75 in pink and riboflavin-free DMEM in lilac. The bracket at later time points indicates increasing cell death. Data from one representative assay are shown.

Figure 4.33 highlights that conditioning these knockouts caused a similar restriction in growth to the commercial SLCKO, Figure 4.6, indicating increased stress due to low riboflavin media. Figure 4.30 shows the OCR profile for each knockout compared to the control, with considerably lower basal oxygen consumption. The results are not as striking as the difference between the SLCKO and control (Figure 4.8) but there was a clear difference between the two knockouts and the parental line – CRISPR KO-1 had a lowered basal rate, ~60%, and CRISPR KO-2, ~50%, of the control. Unlike SLCKO, these cells displayed a decreased OCR with oligomycin addition and recovery, along with some uncoupling in CRISPR KO-1, on FCCP, Figure 4.34. Therefore to establish the extend of any differences between the commercial and homemade CRISPR knockout cell lines, full characterisation of CRISPR KO-1 and -2 is now required to determine whether these lines also manifest a severe OXPHOS phenotype.

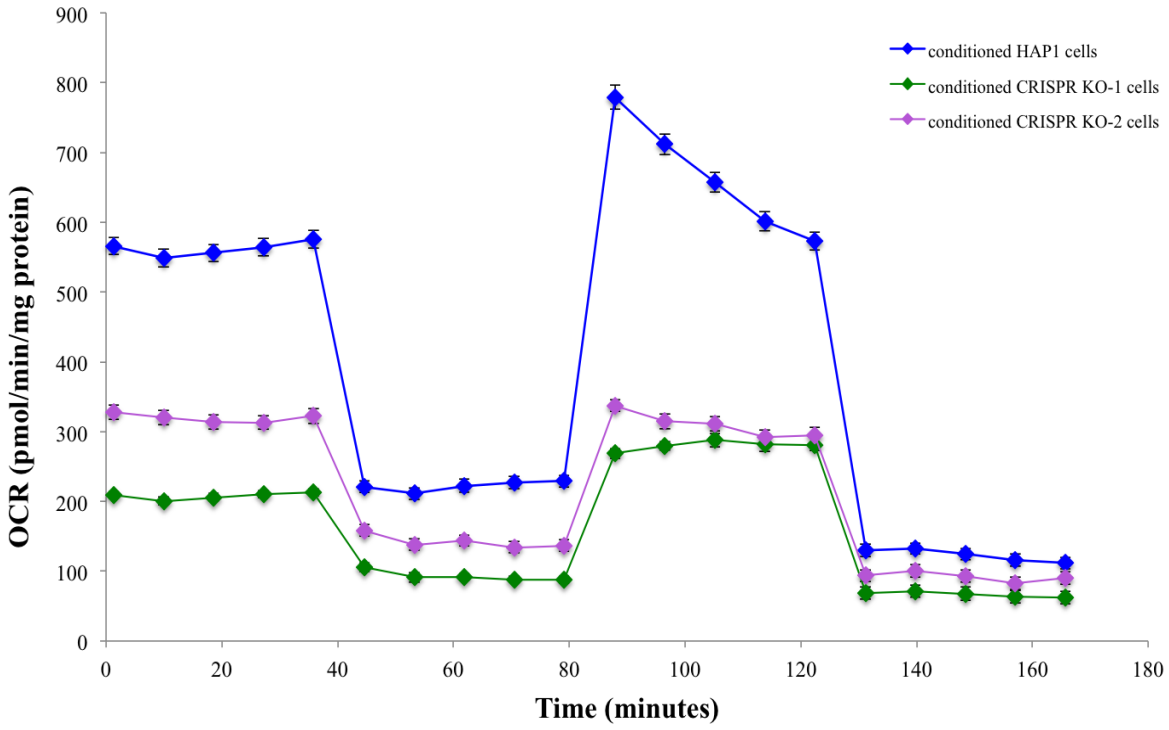


Figure 4.34 Seahorse extracellular flux data comparing the respiratory profile of CRISPR KO-1 and -2 with the parental HAP1 line. Cell lines were conditioned and seeded at 11,500 per well the day before the assay. 32 wells of each condition were assayed and the experiment repeated thrice (mean results for one representative assay shown, with error bars denoting standard error). OCR was corrected for seeding density using BCA protein quantification of wells after the assay. The parental HAP1 line is shown in blue, CRISPR KO-1 in green and CRISPR KO-2 in purple.

4.10. Overexpression of SLC25A32 in knockout cells for phenotype rescue

Another important validation involves the overexpression of the carrier in knockout cells, to show that functional SLC25A32 rescues the phenotype observed. Future work would involve transfection of SLCKO cells for stable expression of the SLC25A32 gene, introduced on a plasmid by lentivirus. A tag would be added to the C-terminus of the protein to allow expression to be checked and also purification of the protein if desired. For this study a small FLAG tag was chosen and inserted before the stop codon of the SLC25A32 sequence. Another component of the transfection plasmid is a selective marker, that can be used to maintain only successfully transfected cells. HAP1 cells are sensitive to the antibiotic puromycin at 1 μg/ml and therefore culturing in puromycin containing media would allow plasmid-containing cells to survive, but remove any that were not transfected. Once a stable SLC25A32 knockout cell line overexpressing the FLAG-tagged carrier has been generated it can then be analysed to determine whether the phenotype has been reversed by the expression of the carrier. This project is now underway with cloning of the plasmid to contain the tagged gene, with a view to taking this forward for expression in the knockout cell lines.

4.11. Discussion

Work presented in this chapter demonstrates that knocking out the previously assigned folate transporter (SLC25A32) had a considerable impact on the respiratory capacity of the cells. Oxygen consumption analysis and in-gel activity assays showed that OXPHOS was decreased and no complex I or II activity was detected. The main aim of this work was to investigate flavin starvation and its effect on complex I, therefore the knockout cell line was used to pursue this aim as the phenotype suggested that SLC25A32 knockout prevented necessary transport of flavin into the mitochondria, starving complex I and II of their cofactor.

SILAC data highlighted the impact on the mitoflavoproteome, SLC25A32 knockout decreased a subset of mitoflavoproteins and some OXPHOS complex subunits, indicating that the flux of flavin into the mitochondria was further restricted, on top of conditioning alone. However, as discussion in section 3.12.2., the use of dFBS may have exacerbated the proteomic effects. Thus, although SLC25A32 cells displayed a strong phenotype, the difficulty of growing them under SILAC conditions may reflect this further decrease in cytosolic riboflavin, affecting their viability. Despite this, it was unexpected that there were no changes to the abundance of complex I subunits when compared to conditioned cells, because data from Chapter 3 suggested that very low flavin caused a decrease in N-domain subunits – thus it would be anticipated that SLC25A32 knockout cells, with a distinct respiratory phenotype attributed to severely depleted levels of mitochondrial flavin, would show a further decrease in N-domain subunits. It is difficult to rationalise why all complex I subunits appeared unchanged; consequently it highlights the importance of characterising more SLC25A32 KO cell lines to determine

whether this effect is substantiated. Complex II subunit levels, in contrast, were lower compared to the conditioned cells, possibly indicating that these proteins are more sensitive to flavin concentrations. Complex III subunits were also decreased, which may be indirectly caused by lack of complex I and II activity.

Transcriptomics explored whether RNA levels were altered and showed that, overall, changes to mitoflavoproteins were likely to result from increased protein turnover due to lack of flavin, rather than from changes to transcription, which would suggest gene regulation. Only complex III showed both down-regulation of subunit expression and decreased protein abundance, so this could indicate a feedback response from the cell to lower levels of complex III.

Complexomics provided further insight into how the structure of respiratory complexes was affected. The data suggest that complex II is not formed in the knockout, whereas complex I is assembled but without the FMN cofactor. Taken altogether these data strongly support the role of SLC25A32 in mitochondrial flavin transport and they also suggest the insertion of FMN into complex I may be the final stage in assembly: without it the complete enzyme is present but inactive.

Further work characterising the CRISPR SLC25A32 knockouts that I generated, and also attempting to rescue to phenotype with over expression of the transporter are now required to validate the results obtained, especially in light of the slightly unusal presence of SLC25A32 peptides covering sequence passed the deletion and transcripts present in the SLCKO. Although CRISPR is now a widely used and effective technique for genomic manipulation there is increasing evidence that indels, which introduce PTCs and prevent expression by nonsense-mediated decay of transcripts, may have unintended genetic consequences. Exon-skipping to produce in-frame transcripts by removing the mutated exon have been identified and could therefore lead to synthesis of partial proteins, which may retain some function. Although not fully understood, these mutations may alter conserved regions within exons known as exonic splicing enhancers and silencers, which regulate alternative splicing (Kapahnke *et al*., 2016). This is an important consideration and the results with SLCKO suggest that the cell line is not a complete knockout. SILAC and transcriptome data support the likelihood of some protein expression, although the cells demonstrated a severe respiratory phenotype correlating to SLC25A32 loss-of-function. Therefore another cell line containing the knockout could act as a control for this outcome and an overexpression rescue would ensure that any additional CRISPR effects had not compounded the biochemical results.

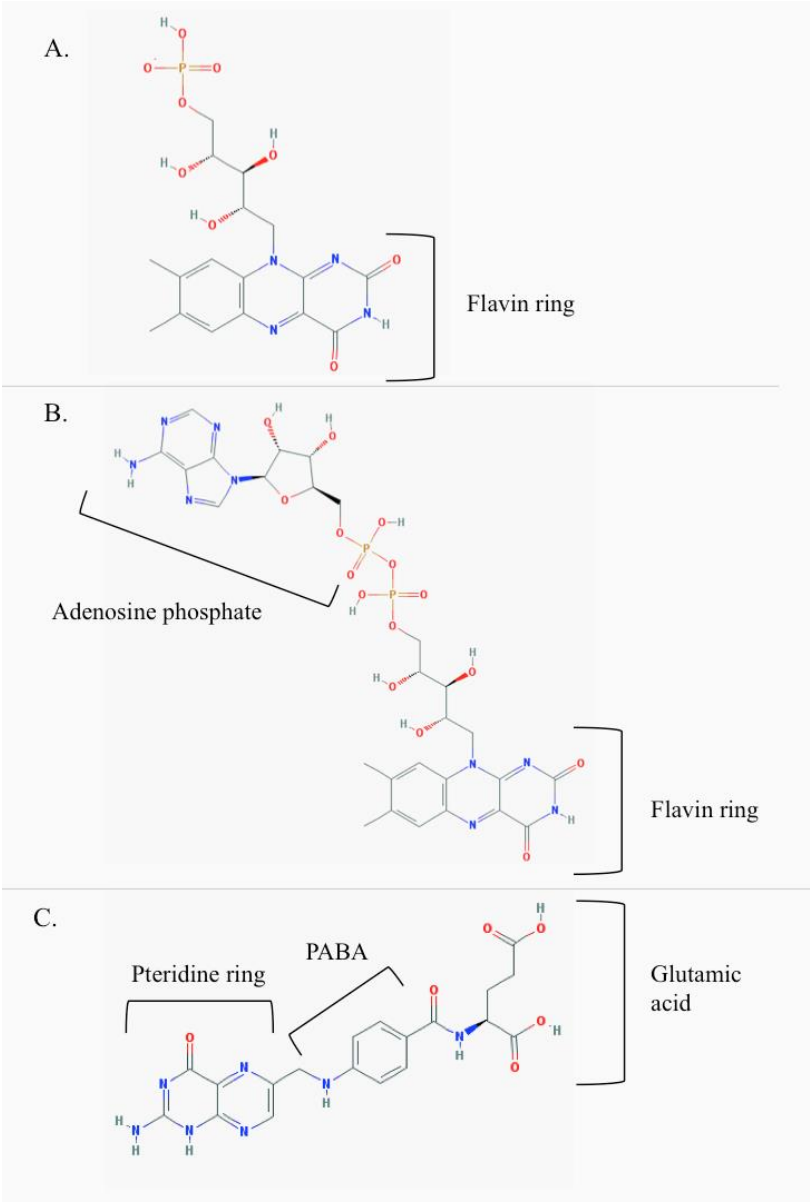


Figure 4.35 Structures of FMN (A.), FAD (B.) and folate (C.). Both flavin species are formed from substituted pteridines with FAD also containing an adenosine phosphate bound to the phosphate of the flavin mononucleotide. Folate is composed of a pteridine ring and glutamic acid moiety connected by 4-aminobenzoicacid (PABA). Structures obtained from NCBI: PubChem Compound Database, 2018.

From the literature the SLC25A32 carrier has two possible functions – as a folate or flavin transporter. As discussed in section 4.1.2. published work has shown rescue by SLC25A32 expression of glycine auxotrophic Chinese hamster ovary-derived glyB cells (folate related phenotype) and a yeast FLX1 null mutant (flavin related phenotype) (Titus *et al*., 2000; Spaan *et al.*, 2005). Two explanations exist for this data: either SLC25A32 is capable of transporting both flavin and folate species or the activity of flavoprotein enzymes in mitochondrial folate metabolism is decreased. The transcriptome data in section 4.7. does indicate some changes in folate handling proteins, with FOLH1 the second most down-regulated gene. However, the data more strongly support the role of SLC25A32 in flavin handling, due to the specific decrease in flavoproteins (particularly NDUFV1 and SDHA). Two patients with mutations in SLC25A32 have recently been reported, interestingly they were riboflavin-responsive and

supplementation let to a marked improvement in their exercise intolerance (Schiff *et al*., 2016*;* Hellebrekers *et al*., 2017). Smeets and co-workers identified a patient with an expected homozygous null mutation in the transporter, therefore for additional riboflavin to be beneficial it suggests that another, potentially low flux, pathway for mitochondrial flavin transport must exist; providing low levels of flavin and can be enhanced by provision of more riboflavin, for uptake and processing in the cytosol. However this has not been explored, but it may explain why SLC25A32 knockouts display severe mitochondrial effects but are still viable because some flavin enters the mitochondria through an independent pathway.

Figure 4.35 shows the structures of FMN, FAD and folate for comparison. It is not known whether FAD and FMN are both transported into the mitochondria; the presence of FADS1 in the matrix suggests that at least FMN would be required, from which FAD could then be generated. Both flavins share the same ring structure, but FAD is much larger due to the additional adenosine moiety, so for both flavin species to be transported SLC25A32 would require flexibility in its binding pocket. Folate does not contain phosphate groups but with a pteridine ring shares some structural similarities with flavins. Additionally, folate and FMN have very similar topological surface areas of 209 and 202 $Å^2$ respectively, suggesting that they would occupy a similar space within a binding pocket – FAD is much larger at 356 $Å^2$ (NCBI: PubChem Compound Database, 2018). This work therefore supports the transport of flavins by SLC25A32 but the transport of folate remains ambiguous, as folate-metabolising phenotypes may be observed due to the importance of flavoproteins in key steps. Therefore, there still remains uncertainty surrounding species transported by the carrier; direct transport assays and greater structural information are required to finally settle the debate surrounding its function.

Chapter 5

Mutagenesis at the ubiquinone binding site in *Yarrowia lipolytica* complex I

5.1. Introduction

5.1.1. *Yarrowia lipolytica* as a model organism

Y. lipolytica is an obligate aerobic yeast that differs from fermentative yeast, such as *S. cerevisiae*, because it preferentially uses respiration and contains an electron transport chain more similar to mammals, including expressing complex I. In addition, its endogenous ubiquinone has a nine-isoprenoid tail (Q_9) - most mammals use Q_{10}, whereas *S. cerevisiae* contains the shorter Q_6 (Kerscher *et al*., 2002). *Y. lipolytica* also contains complex III and IV and expresses an alternative NADH dehydrogenase (NDH2) and alternative oxidase (AOX) as part of its respiratory chain. Figure 5.1 shows that, in addition to complex I, NDH2 carries out NADH oxidation and reduces ubiquinone, but it is a single subunit protein and does not pump protons. Additionally, it is orientated with its active site towards the external face of the IMM. AOX is a terminal oxidase and transfers electrons directly from ubiquinol to water, without pumping protons; it bypasses complexes III and IV, therefore dissipating energy. AOX activity has only been detected in *Y. lipolytica* when complexes III and IV are non-functional, suggesting a role in redox homeostasis (Joseph-Horne *et al*., 2001).

The mitochondrial and nuclear genomes of *Y. lipolytica* have been fully sequenced and annotated, which greatly increases its use as a model system (Kerscher *et al*., 2001; Dujon *et al*., 2004). *Y. lipolytica* has become an established model for complex I structural and functional studies. It has 37 complex I subunit homologues, with eight mammalian supernumerary subunits not found (NDUFA10, NDUFV3, NDUFB1, NDUFB2, NDUFB5, NDUFB6, NDUFC1 and NDUFC2) and four additional components (Angerer *et al*., 2011). A crystal structure at 3.6 Å resolution has been solved (Zickermann *et al*., 2015) and mutations in NDUFS7 and NDUFS8 that cause Leigh syndrome in humans have been explored, demonstrating common changes in V_{max} and binding affinity of hydrophobic inhibitors (Kerscher *et al*., 2002). Additional studies have been carried out on NDUFV1 mutants linked to human pathologies (Varghese *et al*., 2015), along with investigations into the importance of conserved histidines in NDUFS2 (Grgic *et al*., 2004) and the function of particular residues in the quinone-binding site (Tocilescu *et al*., 2010). Therefore *Y. lipolytica* has been shown to be a good genetic model for exploring mechanistic questions of complex I, with the advantages of being a microorganism that can be cultured to produce highly pure and active enzyme.

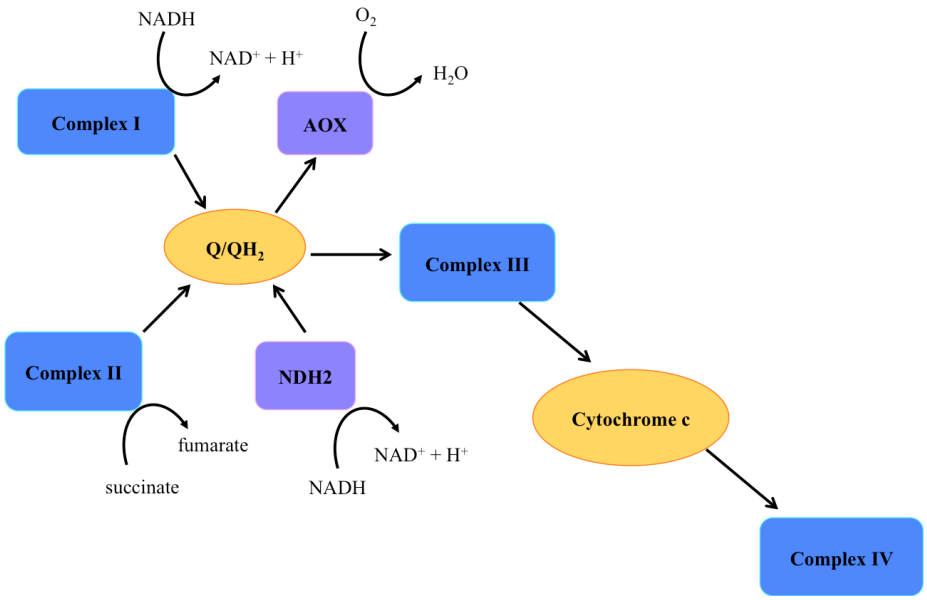


Figure 5.1 The respiratory chain of *Y. lipolytica*. In addition to complex I there is an alternative NADH dehydrogenase (NDH2) and oxidase (AOX). Arrows indicate the direction of electron flow.

5.1.2. The ubiquinone binding site of complex I

The proposed binding site for the ubiquinone head-group in complex I is ~20 Å above the membrane surface, with access through a narrow channel, Figure 5.2A. Although the channel has been structurally identified it has not been definitively proven as the substantial binding channel experimentally. Figure 5.2B illustrates the cavity in the bovine structure, with the head-group binding site found between NDUFS2 and NDUFS7. The residues H59 and Y108 from NDUFS2 are suggested to hydrogen bond the quinone head-group and densities from the *T. thermophilus* structure show this places it within 12 Å of the terminal iron-sulphur cluster N2, for efficient electron transfer (Baradaran *et al.*, 2013). Additionally, the whole channel is long enough to accommodate the entire isoprenoid tail of the substrate, suggesting that this portion of the structure is important. There is a constriction in the tunnel near the entrance, which may influence the reactivity of different substrates and regulate access of inhibitors, Figure 5.2C (Zhu *et al.*, 2016; Fedor *et al.*, 2017).

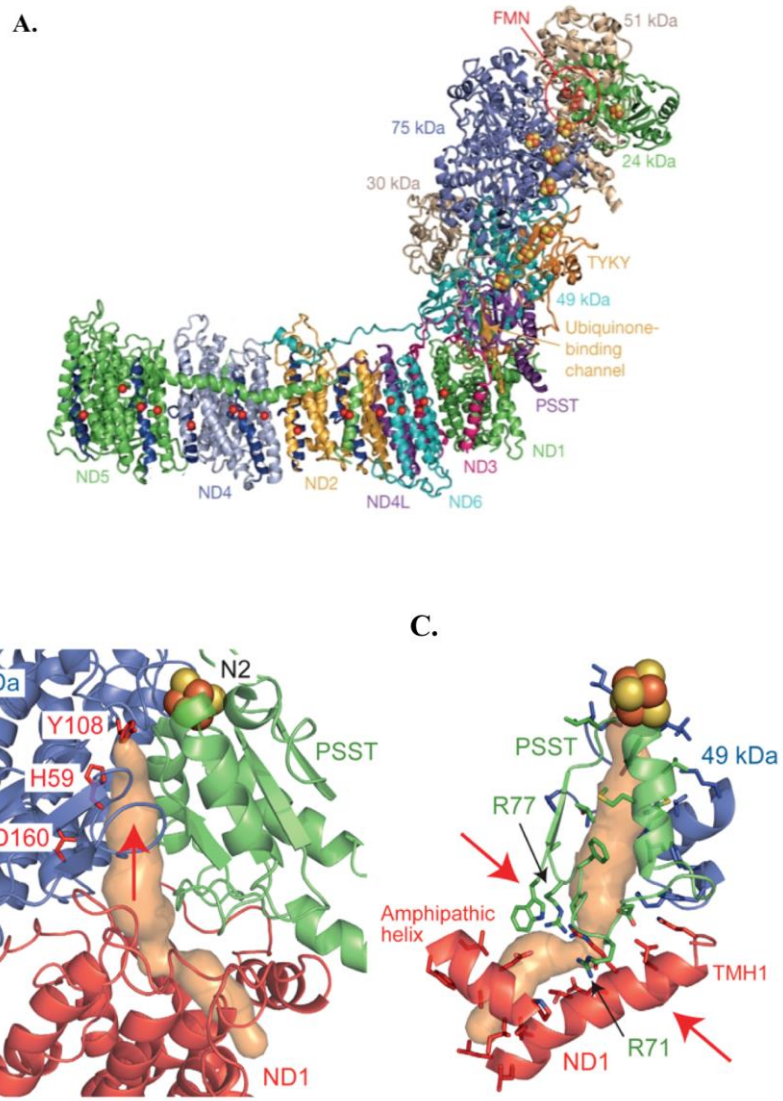


Figure 5.2 The quinone binding cavity of complex I. A. Overview of the core subunits from the bovine structure of complex I and the relative position of the Q-site. The cavity was defined using CAVER software. **B.** The position of the substrate channel is shown in peach, positioned between NDUFS2 (49 kDa, blue) and NDUFS7 (PSST, green) with the entrance formed by ND1 (red). The three key head-group binding residues of NDUFS2 are highlighted and the direction of quinone entry denoted by the arrow. **C.** The bottleneck near the base of the cavity is shown by two red arrows, formed by structural components of the three subunits, including R71 and R77 from NDUFS7. Figure adapted from Zhu *et al*., 2016.

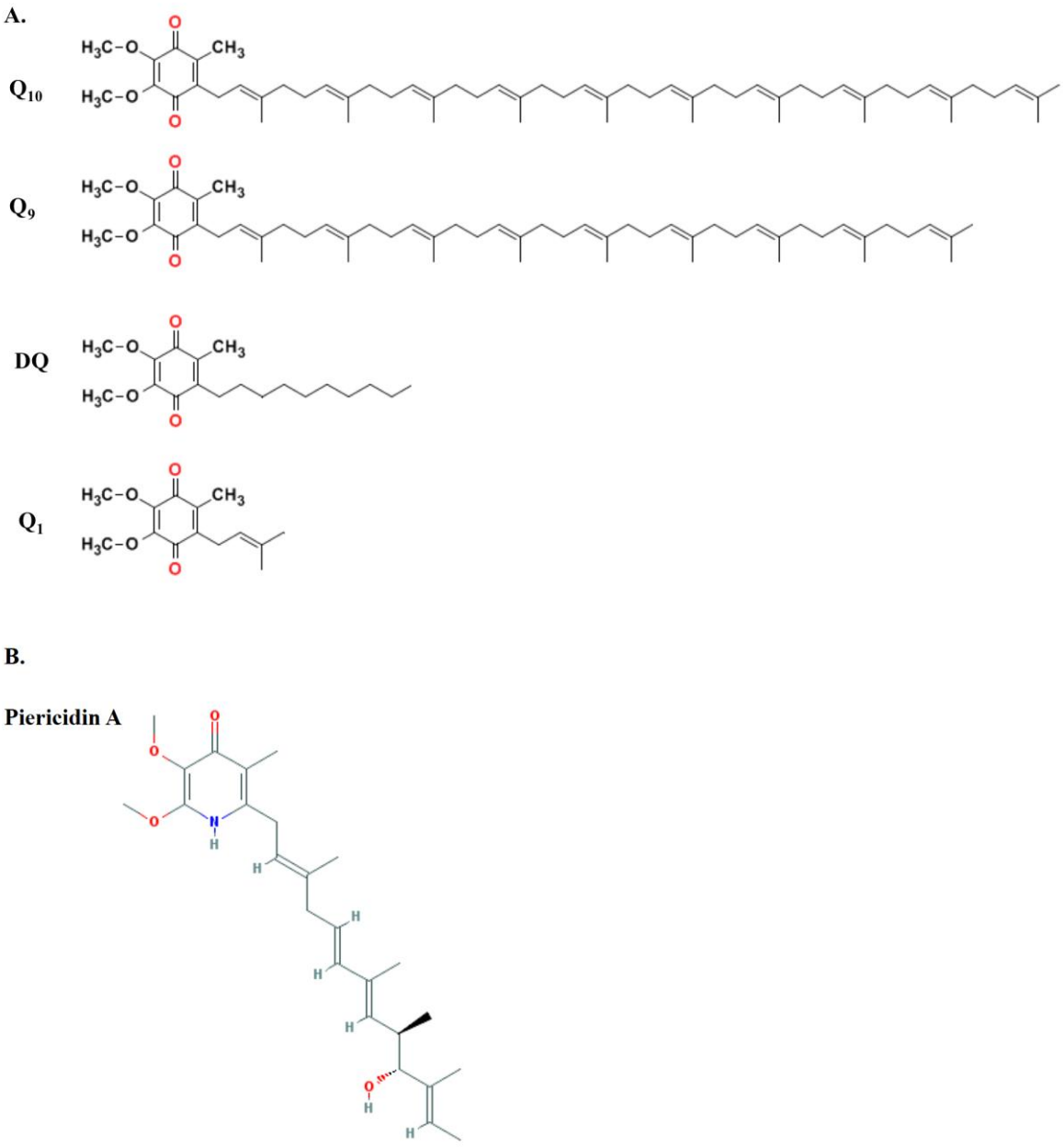


Figure 5.3 Structures of different quinones, and the structurally related inhibitor piericidin A. A. The structures of Q_{10} and Q_9, the endogenous quinones in mammals and *Y. lipolytica* respectively. Structure of the quinone analogue DQ and the smaller more hydrophilic quinone Q_1. **B.** Structure of the structurally related complex I inhibitor piericidin A.

The kinetic properties of complex I and its variants, generated by mutagenesis, are often investigated using non-natural quinone substrates. Long chain quinones, such as Q_9 and Q_{10}, are insoluble in aqueous solution due to their hydrophobicity, therefore a hydrophilic analogue is required – decylubiquinone (DQ). Figure 5.3A shows that, although DQ has the same head-group structure as Q_9 and Q_{10}, its tail is considerably shorter and therefore it behaves more similarly to shorter chain quinones (such as Q_3). The more hydrophilic Q_1, in contrast, exhibits side reactions with complex I, oxidising NADH by a non-proton translocating, inhibitor insensitive pathway. It has previously been shown to interact with the flavin site of complex I, in addition to the Q-channel (King *et al*., 2009). Piericidin A is structurally

related to ubiquinone and a potent inhibitor of complex I (nM range) that was also indicated to bind in the Q-site, Figure 5.3B (Darrozet *et al*., 1998). Therefore binding of this inhibitor provides additional insight into the quinone binding cavity and it is used in complex I activity assays to distinguish NADH:Q activity from NADH oxidation side reactions, such as at the flavin site, which is particularly important when Q_1 is used (King *et al*., 2009). Molecular dynamic simulations investigating the root-mean-square fluctuations of different quinone species indicate that Q_6 and Q_8 behave similarly to Q_{10} and are spatially confined by the quinone binding channel, whereas shorter tail quinones, notably Q_1 and Q_2 show greater conformational freedom (Fedor *et al*., 2017). It would therefore be expected that DQ falls between these two classes (tail length is similar to Q_3) and consequently using different tail length quinones to probe the activity of complex I may provide insight into how mutations affect the structure of the catalytic site.

There is little structural information available concerning the binding of molecules within the proposed Q-site. However Sazanov and co-workers obtained some preliminary data by co-crystallisation of *T. thermophilus* complex I with DQ or piericidin A (Baradaran *et al*., 2013). Difference electron density was calculated prior to ligand modelling. Figure 5.4 suggests that piericidin A, along with the quinone analogue DQ, bind within the proposed channel and their head-groups are positioned similarly, proximal to His59 and Tyr108 of NDUFS2 and ~12 Å from N2. However, the resolution achieved was very low and therefore work is on-going to obtain a more detailed structure showing the positioning of quinone substrates and inhibitors within the proposed Q-site.

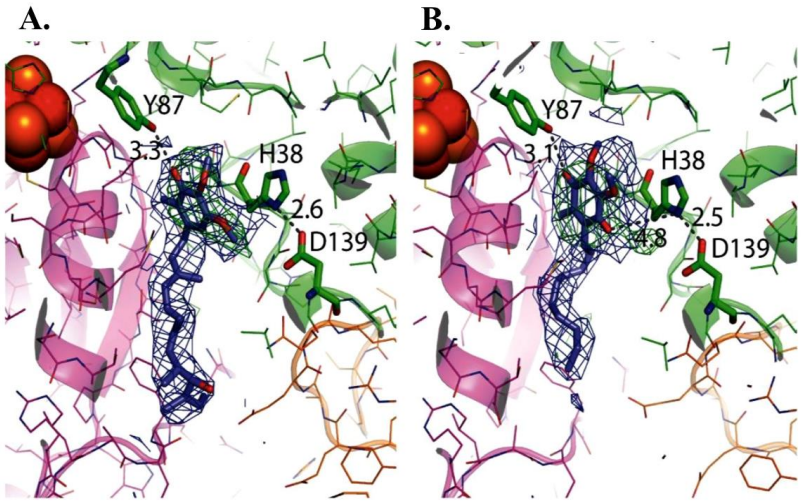


Figure 5.4 Inhibitor and substrate binding in the quinone-binding site of complex I from *T. thermophilus*. **A.** The position of piericidin is shown by experimental electron density and ligand modelling. The three key head-group interacting residues of NDUFS2 (green) are labelled. **B.** Binding of DQ is indicated by experimental electron density and the modelled structure. Residues are numbered for the *T. thermophilus* sequence such that Y87 is Y108, H38 is H59 and D139 is D160 in the human (and bovine) sequence. Figure adapted from Baradaran *et al*., 2013.

5.1.3. Design of mutations in NDUFS2

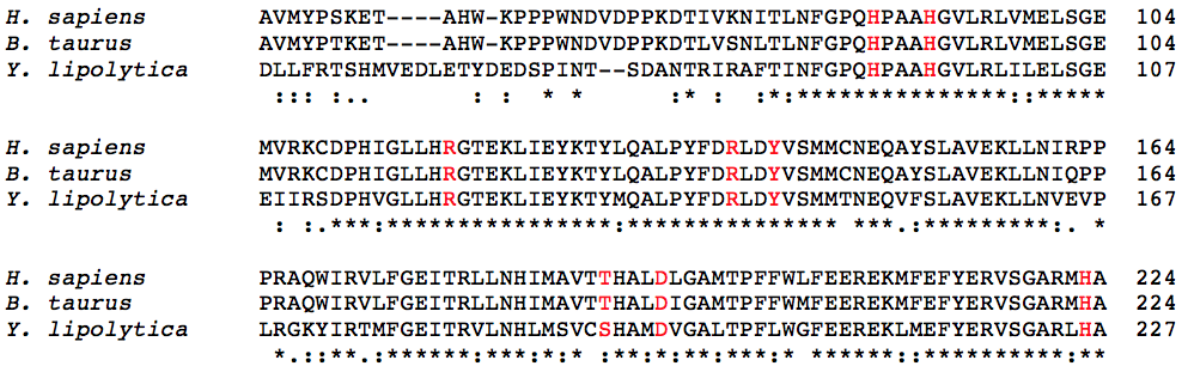


Figure 5.5 Alignment of NDUFS2 sequences from human, bovine and *Y. lipolytica* complex I. A section of the alignment is shown, which contains all the mutations chosen for characterisation in this chapter, highlighted in red. Symbols below denote conservation as analysed by Clustal Omega, where ***** is fully conserved, **:** is strongly conserved, **.** is weakly conserved and a gap demonstrates no conservation.

The core subunit NDUFS2 is 86% conserved across mammals and 63% conserved between humans and *Y. lipolytica*, therefore making mutations in this model system a useful approach for studying the function of specific residues in detail. Point mutations in NDUFS2 were generated by Dr Andrew Jones as detailed in Chapter 2, section 2.9.1., using the Δ*NDUFS2* strain produced by Brandt and co-workers (Grgic *et al*., 2004). This strain was derived from the parental GB10 strain, developed with a His-tagged NDUFS3 subunit and an internally directed NDH2 to compensate for deleterious complex I mutants (Kashani-Poor *et al*., 2001; Kerscher *et al*., 2001). The choice of residues was informed by previous work on the quinone binding site and observations on unusual characteristics of the channel. Figure 5.5 shows the alignment of human, bovine and *Y. lipolytica* NDUFS2 sequences, with the eight residues selected highlighted in red. In this region, conservation between species is particularly high. Figure 5.6 demonstrates the quinone binding cavity of complex I and the involvement of NDUFS2, NDUFS7 and ND1 in formation of the binding site. It also highlights that NDUFS2 is involved in forming part of the upper section of the binding site, interacting with the first three isoprenoid units of the quinone and the terminal iron-sulphur cluster N2. Therefore residues from this subunit have a key role in positioning the head group of quinone and also interacting with the terminal iron sulphur cluster, possibly affecting its redox potential. The structural positions of the mutated residues are shown in Figure 5.7, showing that they form two groups, those found close to the cluster N2 and those facing into the quinone-binding channel. All are conserved across human, bovine and *Y. lipolytica* sequences, apart from S192, which is a threonine in bovine and human complex I.

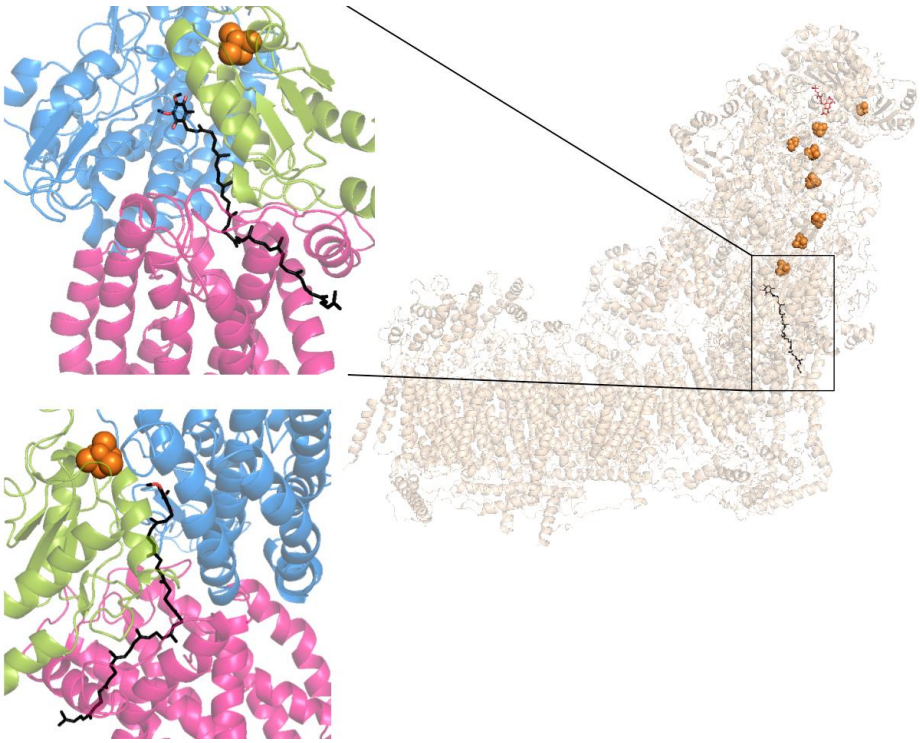


Figure 5.6 Modelled Q_{10} docked into complex I. The bovine structure of complex I (PDB code: 5LC5) showing the position of Q_{10} binding and location of the redox cofactors. Insets show the three subunits involved in the quinone-binding site – NDUFS2 (blue), NDUFS7 (green) and ND1 (pink) and the proximity of Q_{10} to the terminal iron-sulphur cluster N2 (orange spheres). The rotated view shows that the quinone cavity is formed at the interface of NDUFS2 and NDUFS7 with the tail extending down through ND1 to protrude into the lipid bilayer. Structure with modelled Q_{10} from Fedor *et al*., 2017.

Figure 5.7 shows the structural location of the eight residues. Of the five residues chosen that are positioned near quinone in the cavity, three have a relatively well-defined role in quinone binding and reduction. H95 and Y144 (H59 and Y108 respectively in the mature human protein) are both proposed to directly interact with the quinone head-group through hydrogen bonding, protonating the quinone upon its reduction (Sharma *et al*., 2015). Y144 is ~7 Å away from the terminal cluster N2 in the *T. thermophilus* structure and so may provide a connection between the N2 coordinating region and substrate binding pocket (Tocilescu *et al*., 2010). D196 (D160 in the mature human protein) has been suggested to hydrogen bond to H95 (H59 in human), donating protons during quinone reduction (Baradaran *et al*., 2013). The proximity of Y144 to both the quinone head-group and N2 is clearly shown in Figure 5.7, along with the position of H95 and D196. The remaining three residues (R121, R141 and H226) are located away from quinone but near to N2 and may therefore be involved in quinone reduction by modulating its redox potential.

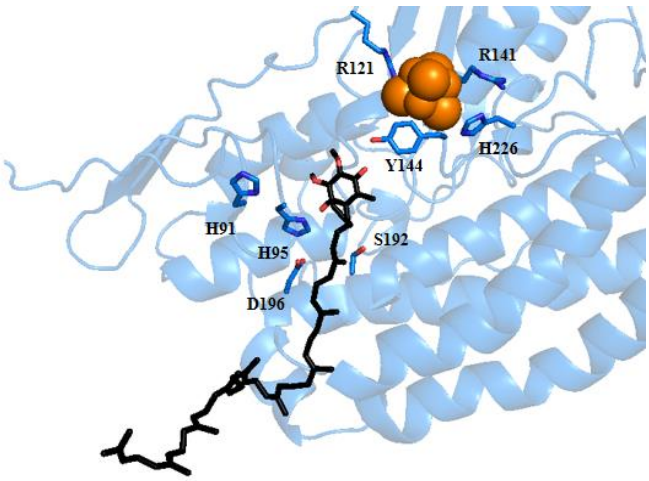


Figure 5.7 Positions of mutated residues and subunit NDUFS2 with Q_{10} docked into complex I. The figure was created using the structure of bovine complex (PDB code: 5LC5) with Q_{10} modelled. Numbering refers to *Y. lipolytica* protein and threonine in the bovine structure at position 156 was mutated to serine to denote S192 in the yeast protein. Q_{10} is shown in black and N2 in orange. Structure with modelled Q_{10} from Fedor *et al*., 2017.

Although there has been previous work on *Y. lipolytica* for some of the NDUFS2 variants chosen, these studies predominantly used mitochondrial membranes, which as an impure system are difficult to characterise and modify. Additionally, for activity assays using mitochondrial membranes it is necessary to use the soluble quinone analogue DQ, rather than the more physiologically relevant long tail quinones (such as Q_9 or Q_{10}), due to their hydrophobicity. Although the reactivity of complex I in mitochondrial membranes with endogenous concentrations of quinone (Q_9 in *Y. lipolytica*) could be measured, by addition of saturating concentrations of cytochrome c (described in section 5.3.2), the concentration of this quinone could not be controlled or modified, making this a suboptimal system for kinetic analysis. The proposed importance of particular residues in quinone binding and reduction instigated further investigation of their kinetic properties, which could be combined with recent high-resolution structural information to gain a better understanding of the channel environment.

To achieve this a complex I proteoliposome system was used, which has been developed by our group. AOX from *Trypanosoma brucei brucei* was modified by removing the N-terminal mitochondrial targeting peptide (residues 1–24) and introducing a Twin-Strep-tag® (Fedor *et al*., 2017). It was then over-expressed in *E. coli* strain FN102/pTbAO. This strain is auxotrophic for 5-aminolevulinic acid as it lacks glutamyl-tRNA reductase (gene *hemA*), which is required for its production in the first step of *E. coli* haem biosynthesis. The inability to produce haem prevents formation of the cytochrome *bd* quinol oxidase and promotes AOX expression (Jones *et al.,* 2016). AOX is incorporated into complex I proteoliposomes to re-oxidise NADH, and the system also allows the inclusion of specified amounts of Q_{10} into the membrane, as shown by the schematic in Figure 5.8 (Fedor *et al*., 2017). This enabled kinetic parameters to be obtained for activity of the variants with Q_{10}. Additionally, the short tail quinone Q_1 and the inhibitor piericidin could be added to assays with proteoliposomes to explore the affect of mutagenesis on their binding. Thus, while previous work informed our selection of NDUFS2 residues

for mutation, using proteoliposomes has many advantages over mitochondrial membranes: components can be added and concentrations precisely defined for kinetic analysis to explore the function of quinone-binding site residues in greater detail.

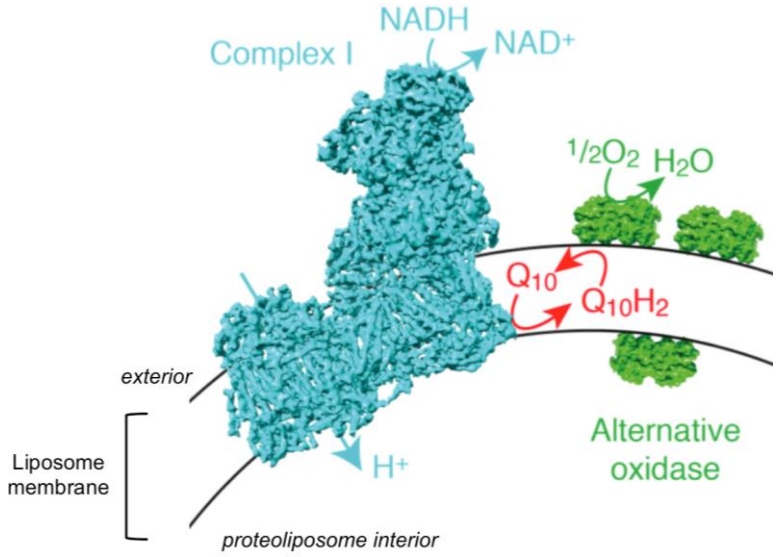


Figure 5.8 Schematic representation of complex I containing proteoliposomes. The proteoliposome system was developed to incorporate defined amounts of phospholipids, complex I, AOX (for quinol oxidation) and Q_{10}, which could all be quantified (Chapter 2, section 2.11). PLs were made for different complex I variants and a range of Q_{10} concentrations were added. These were used for kinetic analysis to obtain a K_M with Q_{10} and in Q_{10} free PLs the addition of exogenous Q_1 allowed the comparison of complex I variant reactivity with different chain length quinones. Figure created by Dr Andrew Jones.

5.1.4. Previous mutagenesis work of NDUFS2 in *Y. lipolytica*

Tyrosine-144

Studies in *Y. lipolytica* by Brandt and co-workers explored the function of Y144, replacing it with six different residues: F, W, I, H, S and R (Tocilescu *et al*., 2010). Activity in mitochondrial membranes showed that for all variants activity was lower than 20% of WT with DQ. However, when the activity with endogenous Q_9 was assayed all mutants demonstrated <5% of WT activity (Y144S and R were not tested). Additionally, the more hydrophilic Q_1 substrate was tested with Y144F/W/H and all demonstrated higher activities, 72%, 93% and 48% of WT, respectively. These results indicate that Y144 is critical for Q-binding and reduction, but very short chain quinones can compensate for the deleterious mutations. Interestingly, EPR measurements on purified enzyme indicated that the type of mutation also affected the redox potential of N2; although Y144F showed no alteration, the mutants Y144W/I/H/S/R all showed a spectral shift or loss of signal. This suggests that Y144 interacts with both the bound quinone and influences the environment of the terminal iron sulphur cluster. To build on this work the Y144F variant was chosen as it represents the most conservative mutation and was found to only have a detrimental effect on quinone reduction, leaving the N2 cluster unaffected.

Histidine-95

This residue is also proposed to bind the quinone head-group and has also been studied by mutation to A, M and R (Grgić, 2004). Activities of H95 mutants in mitochondrial membranes with DQ were all <5% of WT, but EPR determined that N2 was unaffected by the changes. These results indicate that H95 has a key role in quinone reduction through binding of the substrate, but does not influence the properties of the terminal iron sulphur cluster. H95M was chosen for further study where mutation to methionine removes the positive charge but maintains a similar bulkiness of the side chain.

Histidine-91

In addition, three variants of H91 (H55 in human) have been studied - H91A/M/R – they also exhibited activities in membranes with DQ below 5% of WT but no detectable differences in the N2 signal by EPR (Grgić, 2004), suggesting that H91 may also have a key role in quinone binding and reduction. The residue is located on the same loop as H95, which extends into the quinone-binding cavity and may affect quinone positioning or proton transfer, Figure 5.6. H91M was chosen to reflect the histidine to methionine mutation of H95.

Aspartate-196 and Serine-192

There has been no previous mutagenesis work done on D196 in *Y. lipolytica* but its hydrogen bond with H95 is proposed to be important for proton donation, and possibly coupling of quinone reduction to proton pumping (Sharma *et al*., 2015). Therefore, D196N was chosen to hamper the ability of this residue to donate protons to H95. In contrast little is known about the functions of S192 (T156 in human), which is located on the same helix as D196, but further down the channel near the isoprenoid tail of the quinone. S192 is unusual in being a hydrophilic residue within a cavity evolved to bind a hydrophobic substrate, therefore it may be important for hydrogen bonding interactions in the head-group binding region. Although a threonine in human and bovine complex I, serine would retain the same polar properties, albeit with a slightly bulkier side chain. S192V was chosen and non-polar valine used to replace the polar serine.

Histidine-226

H226 (H190 in human) is located near the terminal iron sulphur cluster N2 and proposed to electrostatically interact with the cluster, contributing towards its relatively high redox potential. N2 exhibits a redox-Bohr effect brought about by coupling to H226 (Zwicker *et al*., 2006). Previous mutagenesis work on H226 - with changes to A, Q, C and M - demonstrated that complex I activity with DQ in mitochondrial membranes was decreased by at least 50% in all mutants, except H226M, that

retained 80% activity (Grgić *et al*., 2004). EPR of isolated complex I showed a spectral shift for H226M and decrease in N2 signal height. H226Q showed a considerable decrease in signal height and H226C and H226A demonstrated loss of signal, indicating the importance of this residue on the redox potential of the cluster. Further work showed that H226M abolished the pH dependence of the midpoint potential of N2, which at pH 7 decreased from -140 mV to -220 mV, showing it to be the redox-Bohr group (Zwicker *et al*., 2006). More recently Roessler and co-workers demonstrated that the H226M variant does not impair the redox reaction or proton translocation by complex I and suggested that proton-coupled electron transfer at N2 is not required for catalysis (Le Breton *et al*., 2017). H226M was chosen to further investigate the role of this residue.

Arginine-141

R141 (R105 in human) is also located near N2 and has been studied in mitochondrial membranes as three variants: M, A and K. All demonstrated activities with DQ below 50% of WT; R141A showing the most severe effect with only 17% activity retained (Grgić, 2004). EPR measurements on membranes and purified enzyme demonstrated a marked effect on N2, with loss of signal from all mutants except R141K, which showed 5% of the expected signal height and shifted to lower field values. This effect on N2 supports the role of R141 in tuning the iron sulphur cluster for electron transfer to quinone. To remove the positive charge arginine was mutated to methionine.

Arginine-121

The other arginine proximal to N2 is R121 (R85 in human), which is highly conserved but absent in *T. thermophilus.* EPR data suggests that *T. thermophilus* has an N2 cluster with lower redox potential than NADH, and in *E. coli* it is also lower than other species characterised (Hinchliffe *et al.*, 2006). Interestingly this residue is dimethylated in mammalian, *Pichia pastoris* and *Paracoccus denitrificans* complex I but not in *E. coli*, which provides a direct link between the presence of R121, its methylation status and the reduction potential of N2 (Carroll *et al*., 2013). Therefore R121 was chosen as a residue to study; in line with other arginine residues it was mutated to methionine.

Figure 5.9 shows an in-gel activity assay done by Dr Andrew Jones to check the expression of complex I by flavin site activity staining prior to the characterisation of these mutants. All variants were expressed, although the abundance of some variants appeared lower, notably R121M.

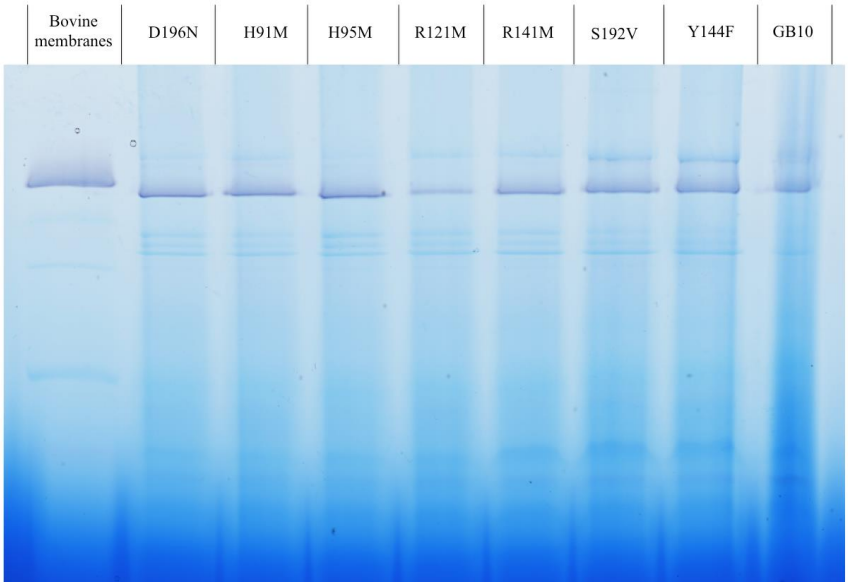


Figure 5.9 BN-PAGE of mitochondrial membranes with in-gel activity staining for complex I. Complex I expression was assessed by Dr Andrew Jones to ensure all mutants expressed complex I with flavin activity before being taken forward for characterisation. Membranes (40 µg) were loaded after solubilisation with 0.6 % DDM. All mutants showed comparable activity to the parental strain GB10 (WT), except R121M that demonstrated ~50%. This indicates that expression levels in the mutants were similar to WT, with the exception of R121M. H226M was assessed previously and also showed comparable activity to WT (data not shown).

5.2. Aims

The aims of this chapter were to:

- Investigate the effects of NDUFS2 variants in *Y. lipolytica* on complex I activity with different soluble quinone substrates in mitochondrial membranes.
- Purify His-tagged variants for isolated complex I activity assays, with the aim of incorporating the complex into proteoliposomes.
- Expand the use of the AOX-containing proteoliposome system from exploring the kinetic properties of complex I to characterise Q-site variants using Q_{10} containing PLs, and with the addition of Q_1.

5.3. Investigating the effect of mutations near the ubiquinone binding site in NDUFS2 on the expression and activity of complex I in mitochondrial membranes

Mitochondrial membranes were prepared from each of the eight variants, along with WT. Kinetic assays were used to gain an overview of the properties of each mutant and the abundance of complex I was analysed by flavin site assays. Data presented in figures are grouped to show: proposed N2-related variants (R121M, R141M and H226M) and those possibly interacting with quinone (H91M, H95M, Y144F, S192V and D196N).

5.3.1. Complex I abundance in NDUFS2 mutant mitochondrial membranes

The abundance of each complex I variant was estimated using two assays, which both measure the amount of flavin present as a proxy for the amount of assembled mature protein. The artificial electron acceptors FeCN and $APAD^+$ react at the flavin site with a ping-pong type mechanism, reacting with the fully reduced flavin in the absence of nucleotide in the binding site (Birrell *et al*., 2009). The main difference is that $APAD^+$ follows a ping-pong reaction, the stoichiometry with NADH is 1:1, whereas FeCN is a one electron acceptor and so this reaction follows a ping-pong-pong mechanism due to the 1:2 NADH:FeCN stoichiometry. It was assumed that none of the mutations at the distant quinone site would influence flavin site activity, therefore these reactions could be used to estimate complex I abundance. Measurements were done in the presence of the quinone site inhibitor piericidin A; so that electrons could not be passed down to quinone as a side reaction.

Figure 5.10 shows the flavin site activity of each mutant compared to WT. Although $APAD^+$ rates showed the largest range (green, Figure 5.10), the trend is similar to the FeCN rates (orange, Figure 5.10). D196N consistently displayed higher activity than WT ($APAD^+$: 192% and FeCN: 128%), whereas R121M showed the lowest activity of all the mutants ($APAD^+$: 75% and FeCN: 74%), which is consistent with the lower observed in-gel activity of R121M shown in Figure 5.7. The data suggests there were some differences in either expression levels or stability of complex I in the variants, R121M and H91M in particular may have deleterious effects on complex I assembly or stability. However, overall, mutations in the Q-site region of NDUFS2 did not affect the abundance of complex I considerably.

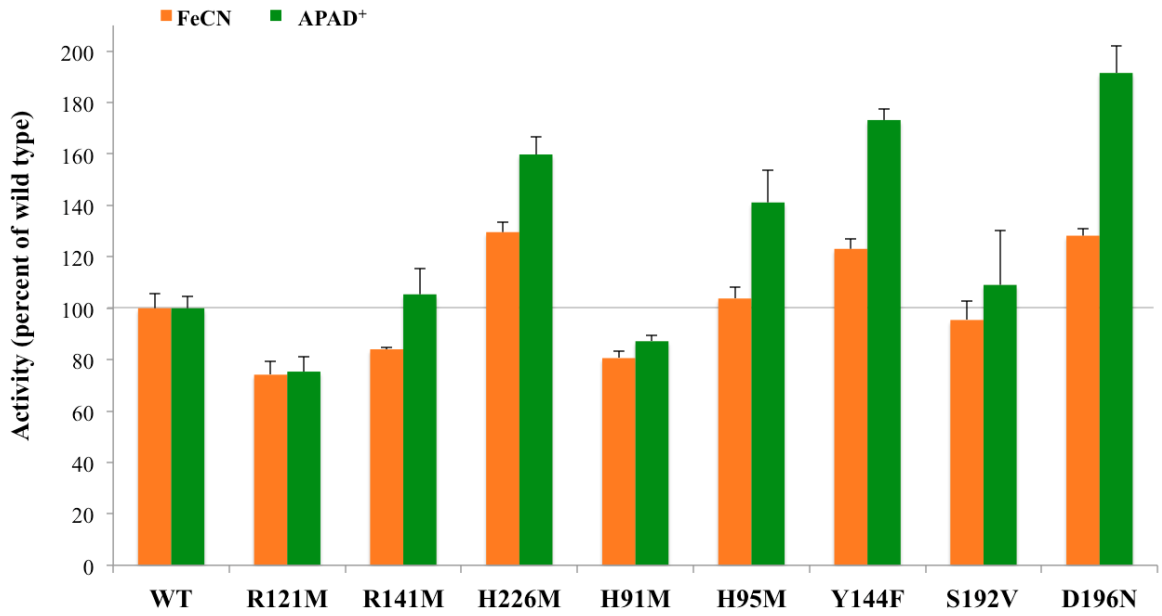


Figure 5.10 Abundance of complex I in variant mitochondrial membranes determined using flavin site assays. Mitochondrial membranes were assayed with 1 mM FeCN and 200 µM NADH in the presence of 2 µM piericidin A, orange bars. Activity was measured by NADH oxidation ($A_{340\text{-}380}$ ε = 4.81 mM^{-1} cm^{-1}) and compared to WT (specific activity: 1.53 µmol/min/mg protein). Mitochondrial membranes were assayed with 500 µM $APAD^+$ and 100 µM dNADH in the presence of 2 µM piericidin A, green bars. Activity was measured by the formation of APADH ($A_{400\text{-}450}$ ε = 3.16 mM^{-1} cm^{-1}) and compared to WT (specific activity: 0.09 µmol/min/mg protein). Each value is the mean of at least six measurements and error bars denote standard deviations.

5.3.2. Complex I activity in NDUFS2 variant mitochondrial membranes differs with different chain length quinones

Figure 5.11 shows the piericidin sensitive activity of each variant with Q_1, DQ and endogenous Q_9. dNADH was used to specifically measure the rate of oxidation by complex I, since NADH would react with other components in the membranes. Overall the eight variants gave substantial decreases in catalytic activity, regardless of the quinone, and all demonstrated less than 50% activity with Q_9 compared to WT. Results are discussed below, although variants that were taken on for further study are considered in more detail later in the chapter.

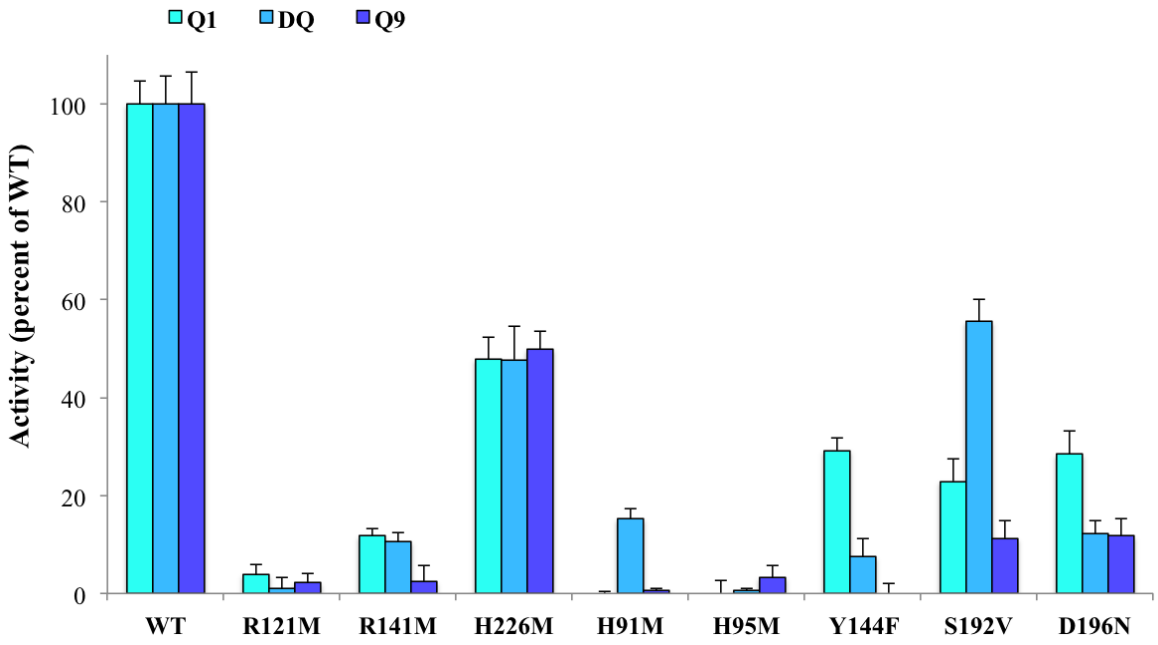


Figure 5.11 Activity of complex I in variant mitochondrial membranes with different chain length quinones. Mitochondrial membranes were assayed with 200 µM dNADH with different quinones. Activity was measured by oxidation of dNADH ($A_{340\text{-}380}$ ε = 4.81 mM^{-1} cm^{-1}). Q_1 was assayed by addition of 100 µM Q_1. Activities were compared to WT (specific activity: 0.57 µmol/min/mg protein). DQ was assayed by addition of 200 µM DQ. Activities were compared to WT (specific activity: 1.53 µmol/min/mg protein). Q_9 was assayed by addition of cytochrome c to a saturating concentration of 8 µM (determined in WT membranes). Activities were compared to WT (specific activity: 0.71 µmol/min/mg protein). All activities are presented as piericidin A sensitive rates (final concentration of 2 µM). Each value is the mean of at least six measurements and error bars denote standard deviations.

The variants R121M, R141M, H95M and H226M all demonstrated a similar level of activity with all three quinone substrates, compared to WT. R121 is located close to N2, mutation to methionine lowered activity to <5% of WT and so catalysis was almost completely abolished. Therefore in *Y. lipolytica* conservation of arginine at this position appears fundamental, distinguishing it from *T. thermophilus* complex I, where the residue is not conserved. The methylation status of R121 was not assessed by Carroll *et al.* in *Y. lipolytica*, but lack of arginine has been correlated with lowered redox potential of N2, which would decrease its ability to transfer electrons to the quinone, due to a smaller difference in their midpoint potentials. This variant was investigated further using the isolated protein (section 5.4). R141 is also close to N2 and activities of the R141M variant were low with all quinones, indicating that it may have a similar function to R121. H95M also showed a similar decrease with all quinone species, despite its suggested interaction with the quinone head-group during reduction. The activities reported here are in line with the measurements by Grgić using DQ in membranes, also <5% of WT; the additional low Q_1 and Q_9 activities suggest the importance of H95 in the catalytic mechanism of quinone reduction and this variant was chosen for further study using the purified enzyme (Grgić, 2004).

Another variant that displayed equivalent rates with all three quinones, compared to WT, was H226M. However this mutation had a milder effect, retaining 45-50% of WT activity. This is surprising given

its proposed function as the Redox-Bohr group for N2 and the methionine mutation was shown to abolish the pH sensitivity of the N2 redox potential (Zwicker *et al*., 2006). Previous work therefore indicates that H226 is close enough to N2 to increase its electron affinity through the influence of its positive charge. However, the mutation to a neutral methionine does not affect complex I activity severely (Le Breton *et al*., 2017).

H91M and S192V demonstrated similar trends in activity, although S192V retained higher activities overall. Figure 5.11 highlights that both variants had highest activity with DQ, compared to WT. The position of both residues within the Q-site indicates that their interaction with the bound quinone would be in the region of the tail, at isoprenoids 1-2, just below the head-group (Figure 5.7). H91 is located on the large loop of NDUFS2 and so this region may be flexible and respond to accommodate the substrate, whereas S192 is found on an alpha helix lining the channel. The possible interactions with isoprenoids near the head-group may account for the differences in activity: in this region Q_9 and Q_1 both share the first isoprenoid moiety, whereas the carbon chain tail of DQ is less bulky, more flexible and may be stabilised differently. The kinetic properties of S192V were explored further using the isolated protein.

The remaining two variants (Y144F and D196N) showed activities that correlated with quinone chain length. Both demonstrated a decrease in activity with increase in chain length: >20% of WT activity was retained with the shortest substrate Q_1 whereas <10% activity was measured with long tail Q_9 (Figure 5.11). Although these mutations are present in the upper region of the binding site, with structural evidence for the interaction of Y144 with the quinone head group (Baradaran *et al*., 2013) and an interaction between D196 and H95 during quinone reduction (Sharma *et al*., 2015), the differing impact of the these variants may relate to the lower predicted conformational freedom of Q_9. As a result Q_9 may be more easily perturbed by mutations, whereas a smaller quinone may have an advantage, by adopting alternative conformations. Therefore both Y144F and D196N demonstrated quinone chain length sensitivity and were investigated further by purification of the variant enzymes.

Table 5.1 summarises the activity data described in this section, along with information from previous studies on *Y. lipolytica* variants. It shows good correlation between activities in membranes with DQ and earlier work; the reactivity with different chain length quinones provides additional insight. These data can be considered with the effect on N2, given by EPR measurements, to further understand how different variants influence complex I activity - either through direct disruption of quinone binding or causing changes to the environment surrounding the iron-sulphur cluster N2, affecting electron transfer to quinone.

Variant	Previous studies			Current work		
	Activity in membranes with (% of WT)	N2 signal EPR measurements	Reference	Activity in mitochondrial membranes (% of WT)		
	DQ			Q_1	DQ	Q_9
R121M	-	-	-	<5%	<5%	<5%
R141M	40%	undetectable	Grgić, 2004	12%	11%	<5%
H226M	80%	decreased intensity and shifted to lower field values, pH dependence abolished	Grgić, 2004 Zwicker *et al*., 2006 Le Breton *et al*., 2017	48%	48%	50%
H91M	<5%	unchanged	Grgić, 2004	<5%	15%	<5%
H95M	<5%	unchanged	Grgić, 2004	<5%	<5%	<5%
Y144F	15% (72% with Q_1, <5% with Q_9)	unchanged	Tocilescu *et al*., 2010	29%	8%	<5%
S192V	-	-	-	23%	56%	11%
D196N	-	-	-	28%	12%	11%

Table 5.1 Summary of data on each of the eight Q-site variants. Previous studies on mitochondrial membranes - activity measurements with DQ and N2 signal monitored by EPR - have been included for comparison with data obtained in this section on variant mitochondrial membranes with different length quinones.

5.4. Purified complex I provides further insight into the properties of five NDUFS2 mutants

From the eight mutants assayed, five were chosen for further study. D196N, H95M and Y144F were chosen due to their suggested roles in head-group binding; R121M was chosen as representative of an N2 interacting residue with a severe effect; S192V was chosen as a less well-understood but potentially interesting residue with different effects with different quinone species.

Complex I was isolated from all variants, although yields were typically lower than obtained from WT; the yield of R121M was only ~50%, which reflects the expression level in Figure 5.9. Complex I from each mutant was then assayed with high DQ (200 μM) and activity compared with WT complex I (specific activity: 45 μmol/min/mg complex I). Figure 5.12 shows that all mutants had measurable quinone reduction rates, although H95M was only ~10% of WT, similar to its low activity in membranes. The other four mutants demonstrated activities of around 30% compared to WT; S192V did not display the highest activity as observed in membranes, Figure 5.11. However using purified complex I in a known reaction mixture provides more robust data than with mitochondrial membranes, which also contain a huge array of other components at unknown concentrations. R121M also demonstrated a differing activity to that measured in membranes, retaining >25%. Since this variant did not show an EPR signal it highlights the requirement to further characterise R121M to better understand changes to catalytic activity.

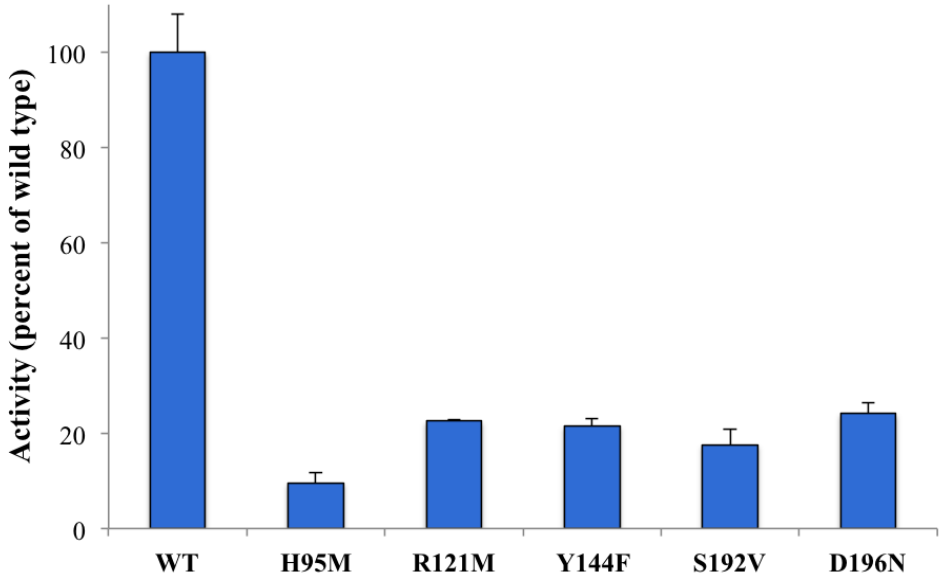


Figure 5.12 Activity of isolated mutant complex I with 200 μM DQ. Purified complex I was assayed at 0.25 μg/ml with 200 μM NADH and 200 μM DQ in 0.075% Asolectin-CHAPS, 20 mM Tris HCl, pH7.5. Activities were measured by NADH oxidation ($A_{340\text{-}380}$, ε = 4.81 mM^{-1} cm^{-1}) and shown as a percentage of the WT (specific activity: 45 μmol/min/mg complex I). Each value is the mean of at least six measurements and error bars denote standard deviations.

In addition to changes in quinone reduction, the effect of mutations on production of reactive oxygen species (ROS) was investigated using the hydrogen peroxide probe Amplex Red (Chapter 2, section 2.10.2). Not all ROS produced is hydrogen peroxide and therefore SOD was used to channel ROS formed towards the Amplex Red reaction (Kussmaul & Hirst, 2006), which is catalysed by HRP and forms resorufin; the fluorescence can be detected by following A_{557} and A_{620} (ε = 51.6 mM^{-1} cm^{-1}). This assay was undertaken in the absence of added quinone. Figure 5.13 shows H_2O_2 production for each mutant compared to WT. R121M showed similar activity to WT, and S192V showed slightly lowered production. Interestingly the three variants implicated in quinone head-group binding displayed an increase in ROS production; Y144F produced ~35% more H_2O_2. ROS is generated by complex I primarily at the flavin site and so this assay can be used to estimate the amount of flavin present (Pryde *et al*., 2011), but it may also provide insights into the different sites of ROS production. The quinone-site has previously been implicated and although this has now been shown as a minor contributor compared to the flavin site, these variants may affect this balance. In the absence of quinone ROS is produced at the flavin site by electrons backing up on the iron sulphur clusters. When quinone is present an increase in ROS production has been observed for piericidin inhibited complex I, by preventing electron transfer to the ubiquinone pool (Fato *et al*., 2009). Changes to the quinone-binding cavity may also affect ROS production, by encouraging electron leak from this site. Therefore it is interesting that while the quinone site is a minor site of ROS generation, quinone-channel variants H95M, Y144F and D196N correlate with an increase in overall ROS production.

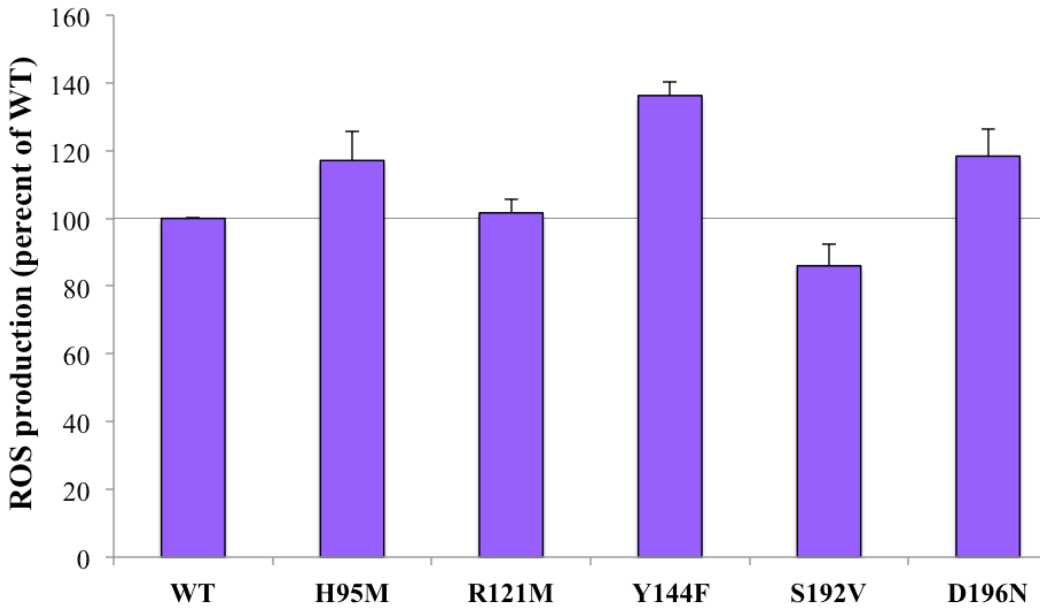


Figure 5.13 ROS production of isolated mutant complex I. Purified complex I was assayed at 10 µg/ml with 30 µM NADH, 10 U/ml HRP, 10 U/ml SOD and 10 µM AxR in 0.075% Asolectin-CHAPS, 20 mM Tris HCl, pH7.5. No protein and catalase controls were also used. Activities were measured by following A_{557} and A_{620}. (ε = 51.6 mM^{-1} cm^{-1} compared with WT (specific activity: 41 nmol/min/mg protein). Each value is the mean of at least six measurements and error bars denote standard deviations.

5.5. Characterisation and kinetic properties of five Q-site variants in proteoliposomes

5.5.1. The proteoliposome system

The proteoliposome (PL) system comprises sealed lipid vesicles containing an 8:1:1 ratio of the phospholipids phosphotidylcholine, phosphotidylethanolamine and cardiolipin. Purified complex I was embedded in the membrane, adopting two orientations – the matrix arm facing the exterior or interior of the PL. Only complex I molecules with the matrix arm facing the outside of the PL will be accessible to added NADH. AOX was also introduced to the PLs during synthesis, to re-oxidase the quinone pool and prevent low concentrations of the quinone substrate from limiting NADH oxidation by complex I. This defined system allows complex I content to be measured and the hydrophobic long tail quinone Q_{10} was added at different concentrations for incorporation into the membrane; the different concentrations could then be assayed to obtain a K_M curve (Jones *et al*., 2016; Fedor *et al*., 2017). Although Q_9 is the endogenous quinone in *Y. lipolytica*, Q_{10} was used instead due to the higher purity of reagent available, and because previous work using bovine complex I in PLs also used this quinone. Due to the similarities in structure and the work done by Fedor *et al.* involving different quinones, the results with Q_9 and Q_{10} were expected to be comparable.

For each of the five variants, and WT, purified complex I was used to produce a series of PLs containing different concentrations of Q_{10}, in addition to proteoliposomes without Q_{10}. Standard complex I-only PLs were also prepared (no AOX or Q_{10}) to provide information for the quantification of complex I present. Each PL type prepared was then split into aliquots to avoid freeze-thaw cycles (as described in Chapter 2, section 2.9.5). The partitioning of Q_{10} into the bilayer was determined using the total concentration of quinone obtained by HPLC, along with the phospholipid content of each of the PLs, therefore kinetic data discussed below is presented as mM of Q_{10} in the membrane. For Q_1 and piericidin additions a similar calculation was done, using published octanol:H_2O partition coefficients and measured phospholipid concentrations for each of the PLs: Q_1 *logP* = 2.9 (Lenaz, 1998) and piericidin A *logP* = 5.02 (Höllerhage *et al*., 2009).

The concentration of complex I present was determined via the flavin site NADH:$APAD^+$ assay, since protein quantification assays would measure both complex I and AOX amounts. Standard PLs containing complex I only were prepared for each variant and WT. The total complex I of these standard PLs was quantified by the Amido Black assay (Chapter 2, section 2.11.2). The NADH:$APAD^+$ activity was determined for these standards and all other PLs (Chapter 2, section 2.11.2). For standard PLs this was done in the presence and absence of 15 µg/ml alamethicin, which permeabilises the membrane to NADH and $APAD^+$, making all complex I accessible to the substrate, regardless of insertion orientation. From this total activity and the Amido Black, which gave the total protein concentration, turnover numbers (k_{cat}) for NADH:$APAD^+$ were calculated for each variant. This value was then applied to the

PLs containing different Q_{10} concentrations and, along with the NADH:APAD$^+$ activities measured for each, it was used to determine the outward-facing complex I present. This was done for all variants; Table 5.2 shows example data for complex I-only standards from WT, D196N, Y144F and H95M, discussed in section 5.4.2. Table 5.3 shows the quantification of the PLs containing seven different Q_{10} concentrations made for these variants (the kinetic analysis of S192V and R121M, discussed in section 5.4.1, used fewer Q_{10} concentrations). The concentration of complex I in each PL was then used to calculate specific NADH oxidation activities in µmol NADH/min/mg complex I, as displayed on graphs below.

	Amido Black assay	**NADH:APAD$^+$ Activity**		**Complex I facing out (%)**	**k_{cat} complex I (s^{-1})**
	Total [CxI] (mg/ml)	**- alamethicin (µM/min)**	**+ alamethicin (µM/min)**		
WT	0.62 ± 0.02	5.5 ± 0.6	9.3 ± 0.7	59.1	168.2 ± 13.5
D196N	0.79 ± 0.01	7.5 ± 0.3	11.4 ± 1.1	65.8	160.6 ± 12.4
Y144F	0.52 ± 0.09	5.2 ± 0.9	7.5 ± 0.3	69.3	126.2 ± 14.3
H95M	0.62 ± 0.05	3.6 ± 0.3	7.6 ± 0.6	47.4	135.9 ± 11.2

Table 5.2 Complex I-only standard proteoliposome data for WT, D196N, Y144F and H95M. These proteoliposomes were used to determine NADH:APAD$^+$ k_{cat} values for each complex I variant and thus to calculate the concentration of complex I facing out in complex I:AOX PLs used in quinone assays, Table 5.3.

WT					
PL type (Q_{10})	**[Phospholipid] (mg/ml)**	**[Q_{10}] (µM)**	**[Q_{10}] (mM, membrane)**	**NADH:APAD$^+$ (µM/min)**	**[Complex I] outward-facing (mg/ml)**
0 nmol	8.5 ± 0.5	9.7 ± 0.5	1.1 ± 0.1	6.8 ± 0.4	0.45 ± 0.04
12.5 nmol	14.9 ± 0.9	16.4 ± 1.0	1.1 ± 0.1	7.4 ± 0.8	0.49 ± 0.07
25 nmol	11.5 ± 0.6	30.2 ± 1.0	2.6 ± 0.2	6.6 ± 0.3	0.43 ± 0.03
50 nmol	11.7 ± 1.4	110.0 ± 6.0	9.4 ± 1.2	7.6 ± 0.3	0.51 ± 0.05
75 nmol	21.0 ± 2.0	141.5 ± 26.5	6.7 ± 1.42	7.5 ± 0.5	0.49 ± 0.06
100 nmol	14.2 ± 0.7	163.4 ± 4.5	11.5 ± 0.7	8.6 ± 0.5	0.57 ± 0.05
150 nmol	15.9 ± 0.9	284.3 ± 16.0	17.9 ± 1.4	9.3 ± 0.4	0.61 ± 0.03

D196N					
PL type (Q_{10})	**[Phospholipid] (mg/ml)**	**[Q_{10}] (µM)**	**[Q_{10}] (mM, membrane)**	**NADH:APAD$^+$ (µM/min)**	**[Complex I] outward-facing (mg/ml)**
0 nmol	22.6 ± 0.6	2.04 ± 3.0	0.1 ± 0.1	6.0 ± 0.4	0.41 ± 0.03
12.5 nmol	14.7 ± 0.6	15.1 ± 2.0	1.0 ± 0.1	7.4 ± 0.5	0.51 ± 0.07
25 nmol	16.3 ± 0.6	29.3 ± 1.5	1.8 ± 0.7	8.1 ± 0.2	0.56 ± 0.07
50 nmol	18.0 ± 0.8	74.3 ± 5.0	4.1 ± 0.3	6.4 ± 0.4	0.44 ± 0.05
75 nmol	19.3 ± 1.2	98.6 ± 10.0	5.1 ± 0.6	5.8 ± 0.3	0.40 ± 0.04
100 nmol	19.0 ± 0.3	152.5 ± 13.0	8.0 ± 0.7	5.7 ± 0.3	0.40 ± 0.06
150 nmol	21.4 ± 1.5	239.4 ± 4.0	11.2 ± 0.8	4.1 ± 0.4	0.28 ± 0.03

Y144F					
PL type (Q_{10})	**[Phospholipid] (mg/ml)**	**[Q_{10}] (µM)**	**[Q_{10}] (mM, membrane)**	**NADH:APAD$^+$ (µM/min)**	**[Complex I] outward-facing (mg/ml)**
0 nmol	19.9 ± 0.9	1.6 ± 2.0	0.08 ± 0.1	7.3 ± 0.5	0.60 ± 0.10
12.5 nmol	29.2 ± 0.9	22.4 ± 2.5	0.75 ± 0.1	8.8 ± 0.7	0.72 ± 0.05
25 nmol	28.5 ± 1.1	48.6 ± 2.0	1.7 ± 0.1	6.3 ± 0.3	0.51 ± 0.03
50 nmol	12.8 ± 0.9	57.5 ± 6.0	4.5 ± 0.6	5.6 ± 0.4	0.48 ± 0.04
75 nmol	19.0 ± 0.6	133.4 ± 0.1	7.0 ± 0.2	4.8 ± 0.4	0.39 ± 0.05
100 nmol	30.6 ± 2.1	165.2 ± 3.5	5.4 ± 0.4	6.9 ± 0.4	0.56 ± 0.03
150 nmol	28.6 ± 1.1	319.9 ±11.5	11.2 ± 0.6	4.8 ± 0.3	0.39 ± 0.10

H95M					
PL type (Q_{10})	**[Phospholipid] (mg/ml)**	**[Q_{10}] (µM)**	**[Q_{10}] (mM, membrane)**	**NADH:APAD$^+$ (µM/min)**	**[Complex I] outward-facing (mg/ml)**
0 nmol	13.1 ± 0.1	2.7 ± 1.5	0.2 ± 0.4	5.1 ± 0.5	0.45 ± 0.05
12.5 nmol	19.5 ± 1.4	15.0 ± 0.5	0.8 ± 0.1	3.9 ± 0.3	0.45 ± 0.05
25 nmol	11.8 ± 0.3	20.3 ± 0.5	2.5 ± 0.1	3.8 ± 0.2	0.34 ± 0.03
50 nmol	24.2 ± 1.3	72.5 ± 2.0	3.0 ± 0.2	2.8 ± 0.1	0.33 ± 0.04
75 nmol	19.3 ± 0.1	89.4 ± 3.0	4.6 ± 0.2	3.2 ± 0.2	0.24 ± 0.02
100 nmol	20.3 ± 1.1	145.2 ± 8.5	7.1 ± 0.6	3.6 ± 0.2	0.27 ± 0.02
150 nmol	16.9 ± 0.85	184.0 ± 3.5	10.9 ± 0.6	4.1 ± 0.3	0.32 ± 0.03

Table 5.3 Quantification of components from proteoliposomes for WT, D196N, Y144F and H95M. Phospholipid concentration and Q10 concentration were used to determine the concentration of Q10 in mM, in the membrane. The activities for APAD$^+$ reduction were used with k_{cat} calculated in Table 5.1 to calculate the concentration of outward-facing complex I in each PL. These data were then used to display activities of quinone reduction in µmol NADH/min/mg complex I and membrane concentrations of Q_{10} as mM, using the phospholipid densities of 1 µl/mg (Nagle & Tristram-Nagle, 2000). Phospholipid data was also used along with published partition coefficients to determine the concentrations of Q_1 and piericidin in the membrane when added to PLs.

5.5.2. Kinetics analysis of the S192V and R121M variants

Proteoliposomes were made for the five mutants selected with a range of Q_{10} concentrations (Chapter 2, section 2.9.5). Additionally a small number of wild type PLs were prepared to compare NADH oxidation rates with Q_{10} and Q_1. This initial data then informed further work and three of the mutants – H95M, Y144F and D196N - were taken forward for more rigorous kinetic analysis (section 5.5.3). Presented below are the initial data for R121M and S192V compared to the WT control, which despite being preliminary, gave an indication of the effect of these mutations on the Q-binding site and the influence of ubiquinone chain length on the rates of NADH oxidation.

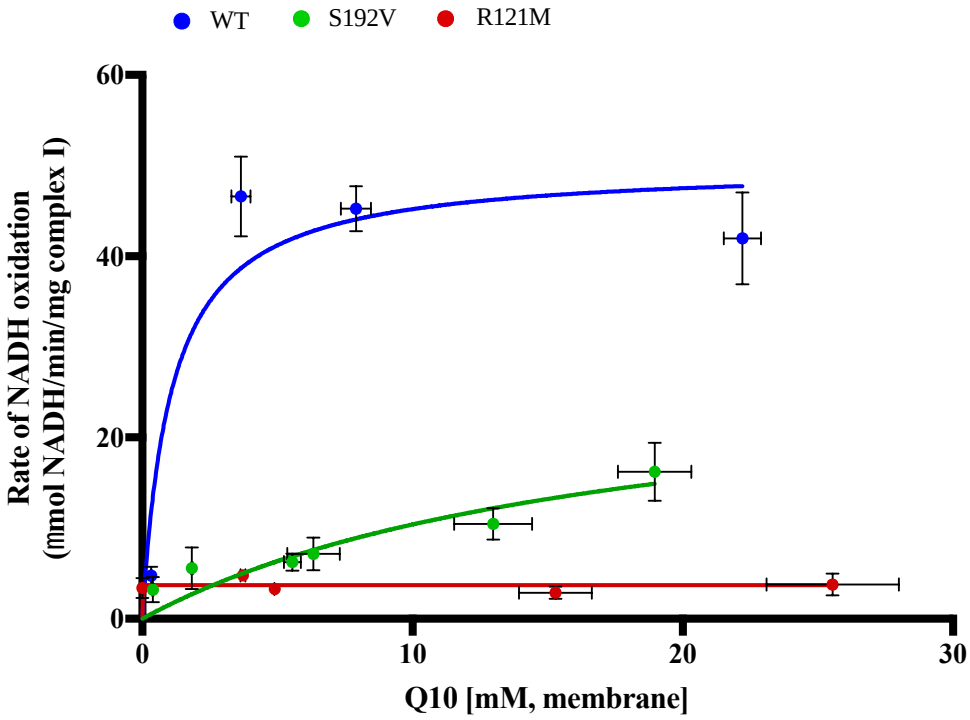


Figure 5.14 Michaelis-Menten plots for NADH oxidation by wild type, S192V and R121M mutant complex I proteoliposomes over a range of Q_{10} concentrations. Proteoliposomes incorporated complex I and AOX with different Q_{10} concentrations. Each value is the mean of at least six measurements; standard deviation was propagated from the underlying measurements. Wild type data is shown in blue, S192V in green and R121M in red.

Figure 5.14 shows initial data on the Michaelis-Menten characteristics of S192V and R121M, which despite relatively large errors demonstrate a considerable change in NADH oxidation rate with Q_{10} compared to WT. Although only four Q_{10} concentrations of WT *Y. lipolytica* PLs were prepared, the data shows that the S192V mutant displayed a largely increased K_M of 17.7 mM Q_{10} (WT K_M of 1 mM Q_{10}) and a lower V_{max} of 28.8 μmol NADH/min/mg complex I (WT V_{max} of 50 μmol NADH/min/mg complex I). This suggested that the change to valine both decreased the maximal rate of NADH oxidation and altered binding of the substrate. In contrast, high concentrations of Q_{10} were unable to rescue activity of R121M; suggesting this position cannot tolerate changes in property, possibly as the positively charged side chain is necessary to mediate electron transfer from N2 to quinone.

To investigate whether the effects of the different variants were confined to long chain quinones, titration with Q_1 was also undertaken. Proteoliposomes without Q_{10} were used and a range of Q_1 concentrations were added to the assay, up to 100 μM - after which partitioning into the membrane is not consistent and may cause destabilisation of the membrane (Fato *et al.*, 1996). The concentration present in the membrane was calculated, as described in section 5.5.1. Additionally, rates were measured with and without the addition of 2 μM piericidin A, to account for Q_1 reactivity at the flavin site.

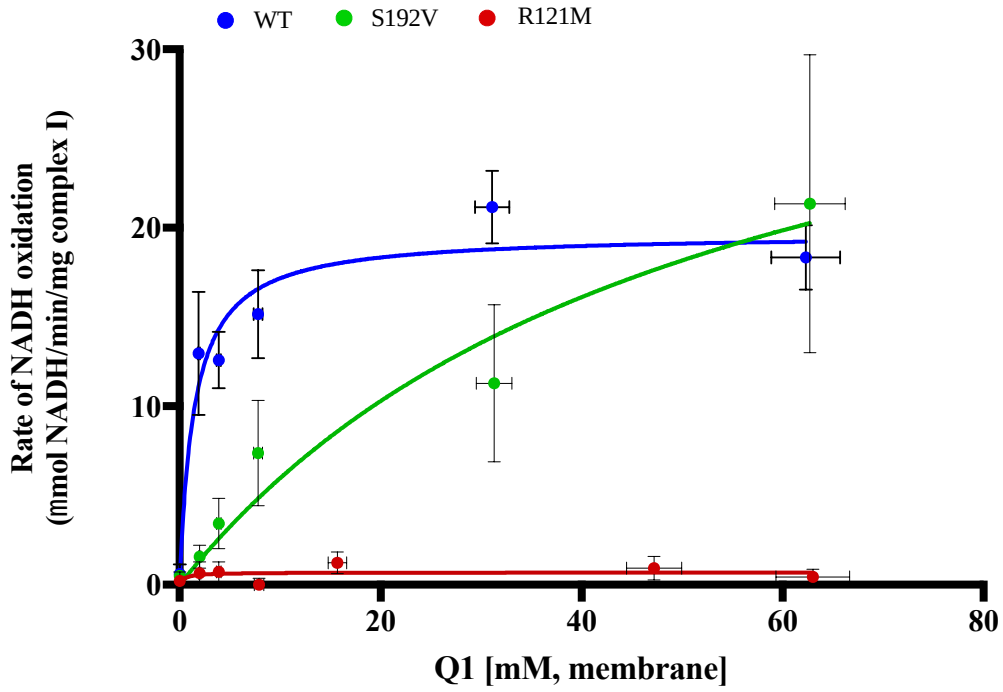


Figure 5.15 Michaelis-Menten plots for NADH oxidation by wild type, S192V and R121M mutant complex I proteoliposomes over a range of Q_1 concentrations. Proteoliposomes incorporated complex I and AOX. Q_1 was added to the assay and concentration partitioned into the membrane calculated. Each value is the mean of at least six measurements; standard deviation was propagated from the underlying measurements. The piericidin insensitive rate (2 μM final concentration) was subtracted. Wild type data is shown in blue, S192V in green and R121M in red.

Figure 5.15 demonstrates the Q_1 dependent NADH oxidation of WT, S192V and R121M complex I in proteoliposomes, kinetic parameters are given in Table 5.4. It is apparent that for WT complex I with Q_1 there was a decrease in V_{max} to 20 compared to 50 μmol NADH/min/mg complex I with Q_{10} – although a similar K_M was maintained (1.5 mM). Fedor *et al.* have studied the effect of quinone chain length on complex I activity from bovine heart and demonstrated that low V_{max} values were obtained with the short chain quinones Q_1 and Q_2 compared to longer chain quinones. K_M values calculated displayed a bell shaped curve, with the highest concentration of Q_4 required and the K_M calculated for long chain quinones Q_8 and Q_{10} was ≤1 mM (Fedor *et al*., 2017). This analysis suggests alteration of the rate-limiting step between Q_1 and Q_2 catalysis compared to Q_4-Q_{10}, as binding and dissociation rates increased with chain length. Modelling also showed that the isoprenoid tails of Q_1-Q_4 do not overlap the comparable units in Q_{10}, demonstrating conformational mobility of these shorter chain substrates in the Q-site. The lack of anchoring or guide function by the tail is proposed to result in lower reactivity of these quinones compared to the endogenous substrate. The Q-binding channel of complex I has evolved to optimise the binding and catalysis of long chain quinones, Q_{10} in humans and bovine and Q_9 in *Y. lipolytica*, although quinones of different lengths share the properties of the head group and can still be reduced. However, structure of the channel is not designed to orientate these smaller substrates in the best position for interaction with catalytic residues or facilitate rapid exchange.

Comparison of kinetic parameters for the two variants with Q_1 and Q_{10} provides insight into how these mutations alter the Q-site (Table 5.4). The V_{max} of S192V with Q_{10} was nearly half that of WT (29 compared to 50 μmol NADH/min/mg complex I), whereas with Q_1 the V_{max} increased to 37 μmol NADH/min/mg complex I and exceeded that of WT (20 μmol NADH/min/mg complex I). The location of S192 in the upper region of the channel, near the first few isoprenoid units following the quinone head-group, may affect Q_{10} orientation that facilitates optimal interactions with catalytic residues; therefore WT maximal rates could not be achieved. However, the altered catalytic effect with Q_1 suggests that maximal rate of NADH oxidation with this quinone was increased; the shorter chain quinone was impacted less by the mutation on the upper Q-channel structure due to the conformational flexibility indicated by Fedor *et al*. This reflects the trend observed in mitochondrial membranes (Figure 5.11) with S192V demonstrating the lowest percentage of WT activity with long chain Q_9, compared to Q_1 and DQ. Thus, despite mutation of this residue, activity was retained with both quinones, which supports the function of S192 as important to the structure of the Q-channel but not catalytically fundamental.

R121 is positioned away from the Q-binding pocket but proposed to interact with the terminal iron sulphur cluster. In EPR studies R121M lacks an N2 signal, therefore mutation to a neutral methionine was hypothesised to severely impair catalytic activity by disrupting the midpoint potential of N2. This was highlighted in Figure 5.14 by very low activity with all concentrations of Q_{10} and a similar profile was observed with Q_1, Figure 5.15. This indicates that regardless of substrate this mutation impairs

catalysis and cannot be rescued (kinetic parameters indicating lack of activity are shown in Table 5.4). Therefore R121M is likely to have affected electron transfer to the quinone, rather than be involved in quinone tail binding. This is in line with observation in membranes and further strengthens the hypothesis that this residue directly impacts the N2 redox potential, rather than participating in the efficiency of downstream quinone binding.

	Q_{10}		**Q_1**	
	V_{max}	**K_M**	**V_{max}**	**K_M**
	(µmol NADH $min^{-1}mg^{-1}$)	**(mM, membrane)**	**(µmol NADH $min^{-1}mg^{-1}$)**	**(mM, membrane)**
WT	50.1 ± 8.3	1.1 ± 1.0	19.7 ± 1.5	1.5 ± 0.6
S192V	28.8 ± 15.6	17.7 ± 16.3	37.1 ± 13.8	52.1 ± 35.9
R121M	3.7 ± 1.9	<0.01	0.7 ± 0.3	0.4 ± 2.0

Table 5.4 Kinetic parameters for WT and NDUFS2 variants in proteoliposomes. Summary table showing V_{max} and K_M with Q_{10} and Q_1 in the proteoliposome system for WT, S192V and R121M complex I.

5.5.3. Kinetics analysis of the H95M, Y144F and D196N variants

As described above a primary set of data was obtained from proteoliposomes from the five mutants, and although this data provided initial results showing changes in Q_{10} and Q_1 reactivity, it also allowed preparation and quantifications to be refined, to minimise measured errors. Therefore a further set of proteoliposomes were made and assessed, along with a better characterisation of WT complex I. For this work a subset of mutations were chosen, those directly associated with quinone head-group binding – H95M, Y144F and D196N.

Figure 5.16A shows the piericidin sensitive activity of the three mutants compared to WT at varying concentrations of Q_{10}. WT demonstrated a V_{max} of 53.8 ± 1.6 µmol NADH/min/mg complex I and K_M of 3.8 ± 0.3 mM Q_{10}, which is consistent with previously published data for the K_M for bovine complex I proteoliposomes with Q_{10} of 3.9 mM (Jones *et al*., 2016; Fedor *et al*., 2017). In contrast, both D196N and Y144F demonstrated decreased V_{max} and a slight shift in K_M, whereas the H95M displayed no observable rate with Q_{10}, regardless of concentration. Figure 5.16B shows the NADH oxidation data over a range of Q_{10} concentrations for D196N and Y144F, to provide clarity due to the scale of Figure 5.16A. D196N demonstrated a V_{max} of 9.1 ± 0.8 µmol NADH/min/mg complex I, which is 80% lower than WT and Y144F also showed a similar trend. There was a marked decrease in V_{max} (6.4 ± 0.6 µmol NADH/min/mg complex I) and slightly increased K_M of 4.3 ± 0.9 mM Q_{10}. Therefore whilst both variants retained activity with Q_{10} the maximal rates were considerably lower. In contrast, H95M displayed no measurable activity and higher concentrations of Q_{10} did not rescue the mutant.

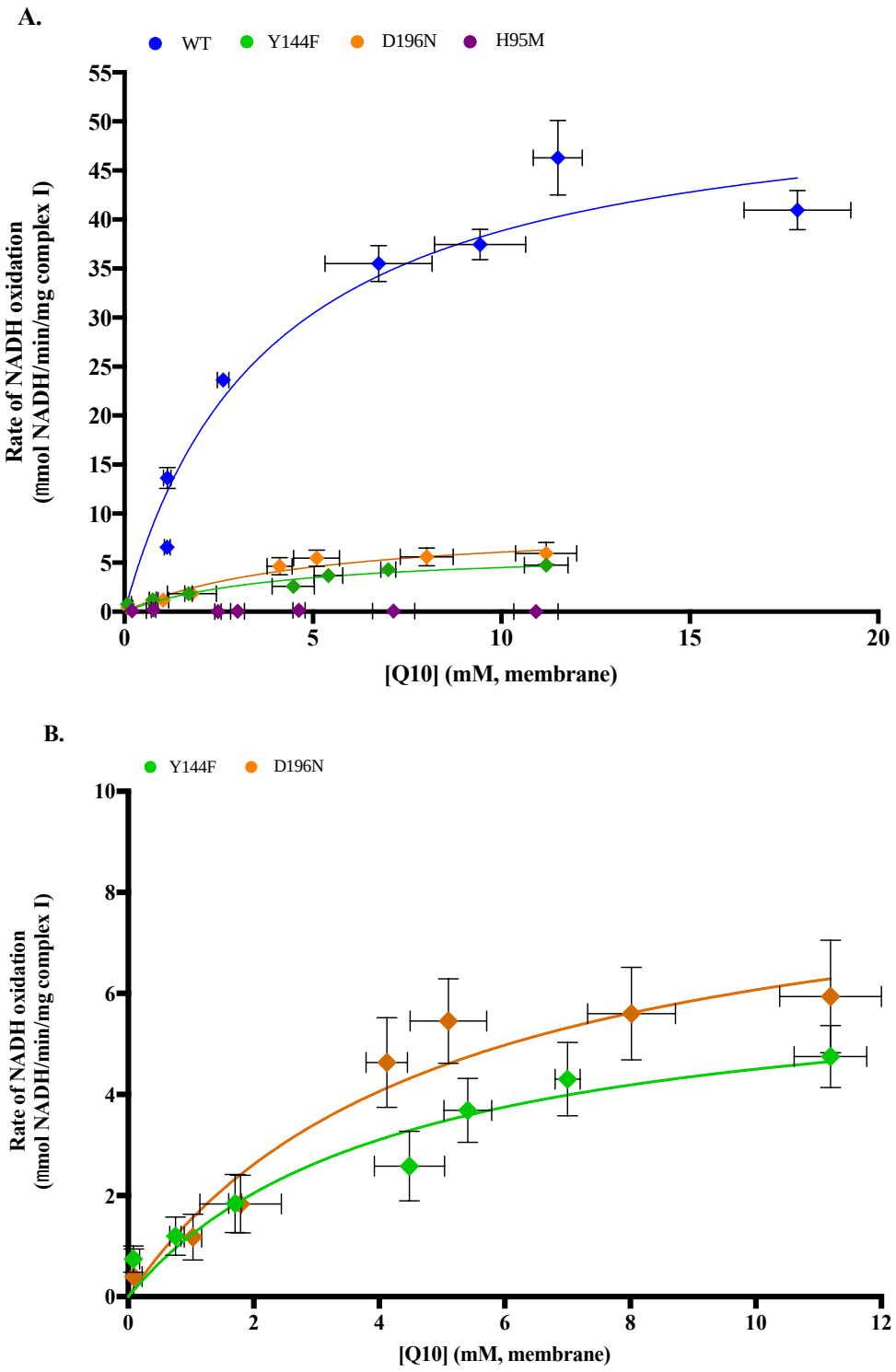


Figure 5.16 Michaelis-Menten plots for NADH oxidation by wild type, Y144F, D196N and H95M mutant complex I proteoliposomes over a range of Q_{10} concentrations. A. Activities for WT, D196N, Y144F and H95M mutant complex I proteoliposomes over a range of Q_{10} concentrations. Proteoliposomes incorporated complex I and AOX and different Q_{10} concentrations. Each value is the mean of at least six measurements; standard deviation was propagated from the underlying measurements. The piericidin insensitive rate (2 µM final concentration) was subtracted. Wild type data is shown in blue, D196N in orange, Y144F in green and H95M in purple. **B.** Zoomed in plot for activities of D196N, Y144F mutant complex I proteoliposomes over a range of Q_{10} concentrations.

Following assessment of NADH oxidation with Q_{10}, titrations were also done with Q_1 - subtracting the piericidin insensitive rate. Once again each mutant showed a considerable decrease in activity, and H95M showed no appreciable catalytic activity, Figure 5.18. Kinetic parameters for WT were calculated: V_{max} 28.8 ± 1.5 µmol NADH/min/mg complex I and K_M 30.6 ± 4.0 mM Q_1, demonstrating a ~50% decrease in V_{max} and substantially higher K_M, indicating the decreased ability of Q_1 to bind and its slower reduction by complex I compared to the longer quinone. The difference in parameters compared to Q_{10} is similar to that measured by Fedor *et al*. in PLs containing bovine complex I (Fedor *et al*., 2017). V_{max} was lowered by two-fold for Q_1 and K_M increased by around eight-fold, demonstrating that although complex I may reduce shorter chain quinone substrates these reactions are much less efficient than with longer chain quinones, for which the binding site has evolved – underpinning the importance of the structure of this cavity for catalysis.

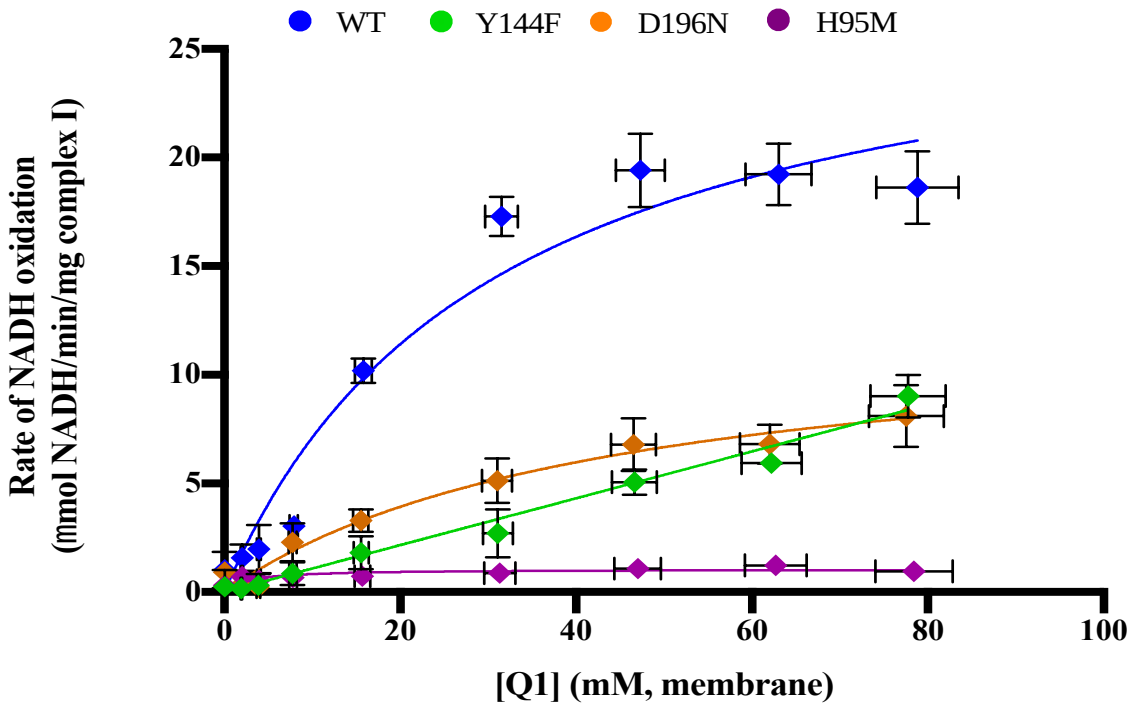


Figure 5.17 Michaelis-Menten plots for piericidin sensitive NADH oxidation of wild type, D196N, Y144F and H95M mutant complex I proteoliposomes over a range of Q_1 concentrations. Proteoliposomes were prepared from isolated complex I with AOX addition and lacking Q10. Q1 was added to the assay and concentration of partitioned Q_1 into the membrane calculated. Each value is the mean of at least six measurements; standard deviation was propagated from the underlying measurements. The piericidin insensitive rate (2 µM final concentration) was subtracted. Wild type data is shown in blue, D196N in orange, Y144F in green and H95M in purple.

Comparison of kinetic parameters obtained for the variants with Q_{10} and Q_1 revealed differences in the effect of the variants on complex I activity. Figure 5.17 shows that D196N demonstrated a slight increase in maximal rate with the shorter tail quinone, but this remained lower than WT. While Q_1 reacted less efficiently with the WT, highlighted by the lower V_{max} and higher K_M, the activity of the mutant was comparable with the two quinones (V_{max} of 9.1 ± 0.8 µmol NADH/min/mg complex I with Q_{10} and 12.5

± 1.2 μmol NADH/min/mg complex I with Q_1) and mirrors the difference in activity measured in the isolated protein with DQ, which was ~20% WT activity. This suggests that with a shorter isoprenoid tail Q_1 achieved a slightly higher maximal rate than Q_{10} for D196N, consequently 43% of WT activity was retained compared to only 17% with Q_{10}. These data reflect the activity measurements in membranes that showed some chain length sensitivity, the highest activity compared to WT was measured with Q_1 (Figure 5.11). This may relate to the altered properties of the head-group binding domain after mutation and that Q_1, with more conformational flexibility, may tolerate this more successfully. Therefore mutation to an asparagine had a dominantly V_{max} effect compared to WT activity, which may encompass changes to the rate of catalysis, expulsion of the product or regeneration of the active site. It is interesting that D196N has this effect because the residue is thought to protonate H95 during catalysis, therefore it would be expected that H95 and D196 variants would have a similar impact. However, although mutation would disrupt this interaction, catalysis may still be possible through an alternative proton donor, which may explain why some activity is retained, whereas H95M is non-functional.

Y144F showed similar kinetic parameters to D196N (Table 5.4) with Q_{10}, however this variant displayed substantial differences with Q_1. The V_{max} with Q_{10} was only ~10% that of WT, whereas the variant displayed ~20% WT activity for rates with DQ measured with isolated enzyme, reflecting the trend demonstrated by membrane activities with different quinones (Figure 5.11); shorter chain quinones retained higher activity and so Y144F appears to display quinone chain sensitivity. This is perhaps surprising due to its equivalent function to H95, interacting with the quinone head group. However, despite the loss of the hydroxyl group, this mutant may retain some activity through adaptions in the Q-binding site. Space may be opened up by the slightly shorter side chain and stabilising interactions may be possible via aromatic stacking with phenylalanine. Although this mutation would prevent hydrogen bonding and proton transfer it may be possible that, as suggested for D196N, another proton donor could be used, due to the slightly altered binding of the head-group, an accommodation more easily made by a flexibly bound short chain quinone. The most surprising aspect of this analysis was that, unusually, Y144F complex I displayed an almost linear trend in activity with Q_1, increasing concentrations resulted in a higher rate, with no indication of saturation. Due to the constraints of partitioning at higher concentrations it was not possible to further increase Q_1 addition to reach the maximal rate and so Michaelis-Menten kinetics could not be applied, only estimated in Table 5.5. However this behaviour suggests a large alteration in quinone binding and further confirms that this mutation has a large effect, with Q_{10} activity considerably decreased and Q_1 appearing to subvert this change and displaying highly changed kinetics compared to WT. Thus the quinone chain length sensitivity of Y144F observed by Brandt and co-workers has been demonstrated in this work; membrane activity decreased with increasing quinone chain length and kinetic characteristic with Q_{10} and Q_1 in proteoliposomes show a distinct change from WT. Therefore although this mutation and D196N have a similar effect on Q_{10}

binding, the effect on the binding of short chain quinones is drastically altered in Y144F; whereas the H95M variant shows complete loss of catalytic ability.

	Q_{10}		Q_1	
	V_{max} (µmol NADH $min^{-1}mg^{-1}$)	K_M (mM, membrane)	V_{max} (µmol NADH $min^{-1}mg^{-1}$)	K_M (mM, membrane)
WT	53.8 ± 1.6	3.8 ± 0.3	28.8 ± 1.5	30.6 ± 4.0
Y144F	6.4 ± 0.6	4.3 ± 0.9	$10<V_{max}\leq 28.8$	$K_M > 50$
D196N	9.1 ± 0.8	4.9 ± 1.0	12.5 ± 1.2	43.7 ± 9.0

Table 5.5 Kinetic parameters for ubiquinone head-group binding mutants of NDUFS2 in proteoliposomes. Summary table showing V_{max} and K_M with Q_{10} and Q_1 in the proteoliposome system for WT, D196N and Y144F complex I. H95M are not shown due to the lack of activity. Y144F showed a linear trend with Q_1, therefore kinetic parameters are estimated given the values obtained and WT parameters.

5.5.4. D196N and Y144F demonstrate altered piericidin sensitivity compared to wild type

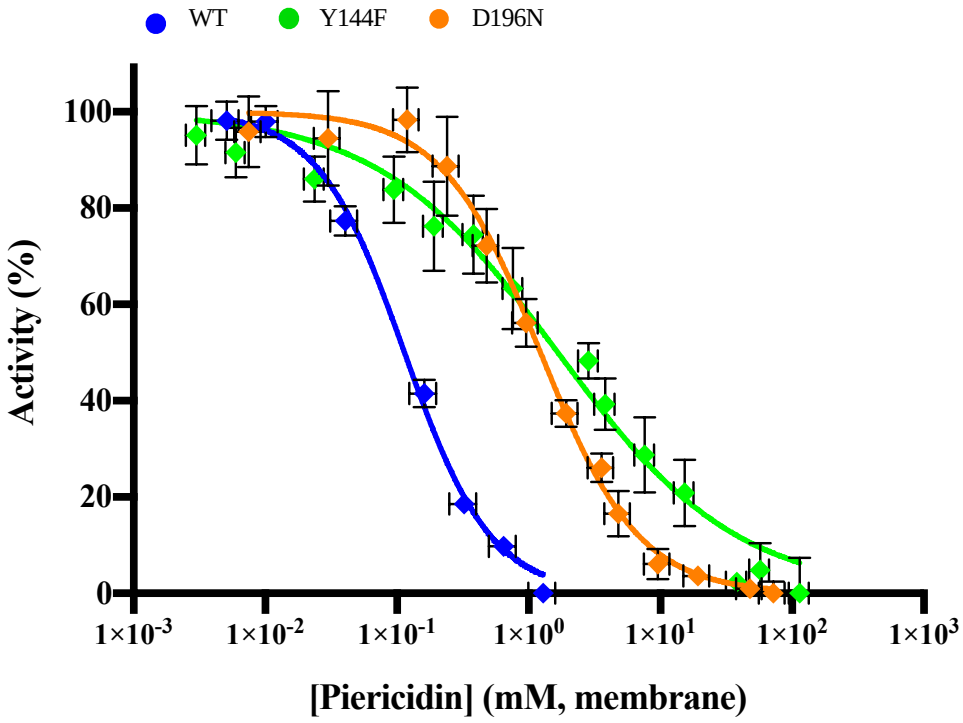


Figure 5.18 Piericidin inhibition plots of NADH oxidation for WT, D196N and Y144F mutant complex I proteoliposomes, at saturating [Q_{10}]. Proteoliposomes were prepared from isolated complex I with AOX addition and a saturating [Q_{10}] chosen - 17.86 mM in WT and 11.2 mM in D196N and Y144F (differences in [Q_{10}] due to sample availability). Piericidin A was added to the assay and membrane concentration of partitioned piericidin calculated. Each value is the mean of at least six measurements; standard deviation was propagated from the underlying measurements. Residual activity was removed from all measurements for a range 0-100% activity. Wild type data is shown in blue, D196N in orange and Y144F in green.

To further characterise the mutations introduced into the head-group binding region of the quinone-binding site the structurally related inhibitor piericidin A was used to determine the sensitivity of D196N and Y144F catalysis, Figure 5.18. Piericidin is a tight binding inhibitor that mimics the enzyme-substrate reaction intermediate and therefore due to the effect of these variants on the Q-binding site it would be expected that they would bind piericidin less tightly. PLs containing a saturating [Q_{10}] in the membrane were chosen and the WT enzyme demonstrated high sensitivity to piericidin with an IC_{50} of 0.24 mM ± 0.05 piericidin in the membrane. The two mutants displayed considerably lower piericidin sensitivity and an altered Hill coefficient. D196N had an IC_{50} of 2.2 mM ± 0.5 piericidin in the membrane and a Hill coefficient of 0.8 compared to the steeper slope (1.5) of WT. Similarly Y144F was also less sensitive to the inhibitor, with an increased IC_{50} of 1.3 mM ± 0.1 piericidin in the membrane. Due to the differences in complex I content in the PLs the IC_{50} values were converted into the ratio of piericidin molecules per orientated complex I for each variant. This resulted in the WT ratio of 6.3 piericidin per complex I in the lipid, matching well with recent data for bovine complex I containing PLs (with 10 mM Q_{10} in the membrane) of ~8 molecules of piericidin per complex I (Fedor *et al*., 2017). D196N and Y144F both demonstrated a large increase, shown in Table 5.6. Y144F required 15 times more piericidin to reach 50% inhibition and D196N was yet less sensitive, the ratio was over 25 times higher than WT. These data correspond well with the observed changes in quinone substrate binding and therefore indicate that piericidin is likely to bind in a similar way to quinone, such that mutations detrimentally affecting quinone reduction also decrease the sensitivity of complex I to the inhibitor.

	Piericidin sensitivity		
	IC_{50} (mM, membrane)	**Hill Slope**	**Piericidin:complex I (in the membrane)**
WT	0.24 ± 0.05	1.5	6.3
Y144F	1.3 ± 0.40	0.6	92.9
D196N	2.2 ± 0.50	0.8	165.9

Table 5.6 Piericidin sensitivity of ubiquinone head-group binding mutants of NDUFS2 in proteoliposomes. Summary table showing the differing kinetic parameters of piericidin binding for D196N and Y144F compared to WT in the proteoliposome system. Piericidin:complex I in the membrane was calculated to account for differences in complex I content of the PLs for each variant. Errors denote standard deviations.

5.6. Conclusions and future perspectives

The work presented in this chapter was designed to build upon current knowledge of the quinone-binding channel of complex I and provide experimental evidence for the importance of residues highlighted in the structures of the enzyme, along with investigating the possible roles of residues found on flexible loops or with unusual hydrophilic properties. Using the *Y. lipolytica* model system allowed genetic manipulation of the NDUFS2 subunit and the engineered Δ*NDUFS2* strain provided an internally directed NDH2 to ensure that mutations compromising catalytic function could still be expressed, due to functional compensation by NDH2. The addition of a His-tag to NDUFS3 also allowed reliable purification and the proteoliposome system was harnessed to provide a defined system that could be used with long and short chain quinones to properly investigate complex I variants.

The more extensive characterisation of five mutants immediately indicated that residues implicated in either quinone binding or terminal iron sulphur cluster stabilisation could both result in loss of catalytic activity when mutated, as exemplified by R121M and H95M, which demonstrated little measurable activity with Q_1 or Q_{10} regardless of concentration. Additionally, some residues likely to interact directly with the quinone head group, and therefore involved in the catalytic mechanism, showed a level of tolerance to mutation; D196N and Y144F displayed altered kinetic parameters and lowered activity with Q_{10}, however with a short chain quinone rates were increased. This apparent sensitivity to the length of the quinone tail provides insight into the relationship between the structure of the channel and the substrate. Although it may not be surprising that residues lining the channel affect the positioning of the quinone tail, this effect was unexpected for residues found in the head-group binding region. As their location precludes them from directly affecting the orientation of the substrate tail, their impact could be linked to the differing behaviour of the substrate in the cavity. The activity measurements with D196N and Y144F indicate that a shorter more flexible quinone may be able to overcome the negative effects of changing these residues and perhaps points to an alternative network of residues, or involvement of water molecules, that are utilised when the integrity of the binding site is compromised. High-resolution structures of complex I with bound substrate or inhibitor in the cavity are still lacking but these two variants may provide useful structural information when compared to WT, particularly if the positioning of Q_1 could be resolved.

Future development of this work could include functional studies in PLs using a wider range of quinones, which could provide further insight into the effects of these particular mutations and the behaviour of substrate based on chain length. Also, with the identification of specific residues an array of mutations could be designed, to gain a better understanding of the residue properties required for function and tolerance at these positions. This work could also be extended to encompass the other nuclear encoded Q-site structural subunit NDUFS7. The altered binding of piericidin A also highlights that D196N and Y144F variants could be used to probe inhibitor binding; it would be informative to test

the sensitivity to other known inhibitors such as rotenone. For other inhibitors that are less well characterised regarding their binding site, this approach could indicate whether these bind at the Q-site or a separate allosteric site.

The advantages of using the proteoliposome system to investigate complex I function and the effects of specific variants also opens this approach up to testing known pathogenic mutations where there is limited knowledge about the mechanism of action. Of the eight mutants investigated in this chapter two of the residues have been linked to complex I deficiency in patients. R121 and R141 were found as R-Q mutations in different patients displaying isolated complex I deficiency and these variants were heterozygous with another NDUFS2 variant (Tuppen *et al*., 2010). Although these patient mutations were further characterised, their presence as heterozygous variants meant that the impact of individual alleles could not assessed. Separation of the variants and assessment in the *Y. lipolytica* system could provide another level of understanding, as both alleles may not contribute equally to the phenotype. The remaining variant residues have not been assigned to disease phenotypes or recorded at high levels in the population as polymorphisms, which may indicate the importance of these amino acids due to lack of variation – embryonic lethal mutations would not be detected. Overall, *Y. lipolytica* represents a valuable model for investigating the function of complex I and has the flexibility to be used for further understanding of complex I mechanistic features by probing specific domains by mutation of nuclear-encoded subunits, in addition to exploring the effects of clinically relevant mutations.

Chapter 6

Exploring the role of high resolution structures in predicting the pathogenicity of human complex I variants

6.1. Introduction

6.1.1. Complex I and disease-causing variants

Complex I is the largest of the ETC enzymes and contains 45 subunits encoded by both the mitochondrial and nuclear genomes. Variation in subunit sequences is linked to mitochondrial disorders with a range of phenotypes and involves both the 14 core subunits and the supernumerary proteins. mtDNA-encoded subunits exhibit the greatest variation, both within and between species – reflecting the faster evolution of the mitochondrial genome (Saccone *et al*., 1999). Although rare, severe mutations cause disease whereas common genetic variants, present in >1% of the population, define haplotypes – the mtDNA background shared by groups of individuals. Despite being considered as benign changes in themselves these differences have been associated with disease, due to the potential to promote the clinical penetrance of other mutations in complex I genes or increase the risk of developing other diseases linked to complex I defects (Dimauro & Davidzon, 2005). Effects on penetrance have been observed in LHON patients, the disease is often caused by the m.14484T>C *ND6* mutation, however increased penetrance is associated with the J haplotype, in contrast to low penetrance of individuals with the H haplotype – suggesting that genetic background has a strong influence on disease manifestation (Howell *et al*., 2003). The clinical presentation of mitochondrial diseases is also linked to the level of mitochondrial heteroplasmy, with different variants demonstrating different thresholds – above which the mutation is pathogenic. The level of heteroplasmy may also differ between tissues and therefore dictates the level of organ involvement in the phenotype (Stewart & Chinnery, 2015). In a similar way, variation in nuclear-encoded genes may result in severe homozygous or compound heterozygous mutations linked to mitochondrial diseases, or heterozygous variants acting as risk factors for disorders. Therefore, a major challenge in mitochondrial medicine is to differentiate between pathogenic mutations and benign polymorphisms in mitochondrial disease patients, to ensure the correct variant is attached to the phenotype – otherwise future use of the information in genetic counselling or research will be severely compromised.

6.1.2. Current tools for analysis of mutations and prediction of clinical outcomes

The evaluation of clinically identified novel genetic variations uses, predominantly, primary sequence analysis – particularly conservation and the severity of the amino acid change incurred. It also includes consideration of changes to the secondary structure and some structural analysis. Many *in silico* tools

have been developed to facilitate this, including PolyPhen2 (Adzhubei *et al*., 2010), SIFT (Ng & Henikoff, 2001) and MutPred2 (Pejaver *et al*., 2017). In a number of studies comparing different tools it has been noted that including structural data improves the quality of the prediction and indicates that consideration of a range of protein features is required (Walters-Sen *et al.*, 2015). However, these tools do not always agree on the predicted outcome and therefore clinical guidelines encourage the use of multiple tools to interpret the outcome of a variant (de la Campa *et al*., 2017). Consequently, lack of agreement on some variants means they cannot be interpreted with confidence. More recently software has been developed to combine results from different tools, to improve reliability. These include the rare exome variant ensemble learner (REVEL) method that collated pathogenicity predictions from 18 tools, eight conservation-based and ten functional-based, to assess the possible outcome of rare missense variants (Ioannidis *et al*., 2016). REVEL was found to outperform other combined features tools to distinguish rare disease from neutral variants and highlights the importance of considering a range of prediction methods. However it still relies on pre-existing tools, which are limited in the features of the residue and protein used by the prediction algorithm. Additionally, despite these tools being useful to prioritise genetic variants, they have not been developed for individual proteins and so the question remains as to whether the sensitivity and specificity of predictions could be increased by consideration of more detailed information on a particular protein (Pejaver *et al.*, 2017).

High confidence predictions are useful in both clinical and research settings, providing a starting point for genetic counselling or possible treatment and informing the investigation of biologically relevant regions of a protein. Thus, computational approaches can highlight residues or domains that are likely to be severely impacted and biochemical assays can then be used to probe them further, to enhance our understanding of structural and mechanistic features of the protein that cause this susceptibility.

6.1.3. Structural information on complex I may improve understanding of pathogenic mutations

Cryo-electron microscopy has revolutionised structural knowledge of mammalian complex I, yielding high-resolution structures with all 45 subunits located and modelled. The first of these was bovine complex I, with all subunits located and modelled onto a 4.2 Å structure in 2016 (Zhu *et al*., 2016). Importantly these structures detail the residue positions within subunits and demonstrate interactions between domains, highlighting functionally relevant regions of the enzyme. Benign polymorphisms and disease-related variants have been identified in human complex I genes and, by comparison of these two variant types, trends may identify potential disease hotspots in the enzyme. Although this approach has been used to provide a general explanation for some variants causing disease, such as those involved in LHON and Leigh syndrome, there has been no detailed global assessment of the high-resolution structures in light of the position of identified human variants (Fiedorczuk & Sazanov, 2018).

Therefore to increase the reliability of pathogenic outcome prediction for complex I variants a human model was generated from the most complete and highest resolution complex I structure. This was solved for active complex I from mouse heart, at 3.3 Å, with the positions of the majority of its 8,500 residues located and 97% of residues mapped with side chains (Agip *et al*., 2018). Human complex I gene variants were collated from online databases, as discussed in section 6.3, and classified as either pathogenic mutations or polymorphisms into three classes for each, based on the confidence of their assignment. These six classes were then analysed in relation to the human complex I model to assess features that have predictive power for the outcome of the variant. This ranged from primary sequence analysis of the conservation across mammals, determined by alignment of ~20 species sequences (Table 6.5), through to secondary structure and finally incorporating the detailed knowledge gained from the high-resolution structure. By considering the unique and functionally relevant features of complex I this analysis aimed to add depth to current prediction tools and enhance the confidence with which predictions about variants across the whole structure can be made.

6.2 Aims

The aim of this chapter was to bring structural data to bear on clinically relevant observations and improve our understanding of complex I variants - specifically to:

- Collate and classify all variants in complex I subunit encoding genes as either pathogenic or polymorphic.
- Analyse these variants on multiple levels: primary sequence, secondary structure elements and three-dimensional structure.
- Draw conclusions about different features and their predictive power.
- Construct a comprehensive table for analysis to provide a prediction for a given residue, which can be used in a clinical setting to improve the confidence of pathogenic mutation assignment.

6.3.Assigning database variants as pathogenic or polymorphic by degree of confidence

A database of 271 residues with likely pathogenic variants (termed mutations) and 608 residues with polymorphic variants (termed polymorphisms) in human nuclear and mtDNA complex I genes was compiled using online resources, Table 6.1 and 6.2. Variants assigned as pathogenic (mutations) in nuclear-encoded complex I genes were extracted from the Human Gene Mutation Database (HGMD, Stenson *et al*., 2017 - http://www.hgmd.cf.ac.uk) and the National Centre of Biotechnology Information (NCBI) ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). HGMD is a curated database that excludes unpublished or inadequately described variants. ClinVar includes a clinical significance descriptor provided by the submitter. Variants were checked to ensure they related to the canonical sequence and not an alternative or predicted transcript. Additionally, variants causing a premature stop codon were removed and only missense/non-synonymous mutations were retained. Mutations were then classified by the confidence level of their assignment. Class 1 mutations (highest confidence) were described as 'pathogenic' by ClinVar AND 'complex I deficiency' causing by HGMD. Class 2 mutations were either denoted 'pathogenic' by ClinVar OR 'complex I deficiency' causing by HGMD. Class 3 mutations (lowest confidence) were rated by ClinVar as 'likely pathogenic' or 'uncertain significance'. Those assigned as 'benign', 'likely benign', 'protective' or 'haplogroup marker' were discounted. For mtDNA-encoded subunits data was taken from ClinVar in the same way and also from MITOMAP (Lott *et al*., 2013 - http://www.mitomap.org). Class 1 mutations were assigned as 'pathogenic' by ClinVar AND 'confirmed' by MITOMAP. Class 2 mutations were either rated 'pathogenic' by ClinVar OR 'confirmed' by MITOMAP. Class 3 mutations were denoted 'likely pathogenic' or 'uncertain significance' by ClinVar and/or 'reported', 'conflicting', 'unclear' by MITOMAP. Where multiple pathogenic variations were identified for the same residue analyses were based on the highest confidence level present. Importantly, the classification does not denote disease severity; it refers only to the confidence of assigning the variant as the cause of a specific inherited mitochondrial disease such as Leigh syndrome or LHON.

Polymorphisms were identified in nuclear-encoded complex I genes using the Exome Aggregation Consortium (ExAc, Lek *et al*., 2016 - http://exac.broadinstitute.org/), with exome sequencing data from 60,706 control individuals, and in mtDNA-encoded subunits using 30,589 mtDNA sequences from MITOMAP. Analyses based on the residue position rather than the identity of the amino acid change combined variants at the same location for a total polymorphic frequency at that position. Class 1 polymorphisms (highest confidence) vary from the canonical sequence with a frequency of >0.01 (>1%), class 2 polymorphisms with >0.001 frequency (>0.1%) and class 3 polymorphisms (lowest confidence) with >0.0001 frequency (0.01%). Variants with frequencies lower than 0.0001 were discounted. Databases were mined by Dr Wei Wei, from the MRC MBU, University of Cambridge, after which I carried out the variant classifications. The features of all variant residues discussed in this chapter are tabulated in Appendix 7.9A-D.

Subunit	Chain	Class 1	Class 2	Class 3
NDUFV1	F	97, 102, 186, 194, 403	36, 68, 91, 127, 179, 184, 191, 226, 232, 321, 357, 358, 366, 412	[7], 19, 75, 131, 133, 140, 247, 263, 297, 337, 348
NDUFV2	E	-	177	153, 158
NDUFS1	G	208, 218, 229, 385, 572, 596	48, 143, 205, 499, 672, 684	39, 154, 181, 230, 286, 408, 432, 452, 483, 511, 543, 553, 577, 680
NDUFS2	D	195, 196, 259, 380, 393	77, 85, 105, 115, 191, 290, 300, 410, 413	20, 287, 347, 376
NDUFS3	C	109, 163	187	210
NDUFS7	B	84	67, 107	77, 81, 167
NDUFS8	I	29, 43, 45, 51, 68, 104, 125	20, 60, 120	71, 81, 109, 127, 170
ND1	H	24, 52, 110, 131, 132, 143, 195, 289	1, 28, 30, 31, 59, 164, 208, 214, 215, 240, 277, 285	2, 4, 11, 12, 25, 34, 38, 39, 56, 64, 113, 128, 144, 147, 176, 187, 218, 230, 276, 279, 304
ND2	N	-	57, 71, 259	55, 60, 64, 122, 128, 150, 328, 331
ND3	A	34, 45, 47	26, 60, 66	10
ND4	M	340	109, 165, 313, 409	158, 206, 372, 387, 412, 423
ND4L	K	65	32	25, 41, 71
ND5	L	124, 393	1, 145, 159, 171, 236, 237, 250, 434, 465	9, 21, 96, 99, 100, 149, 239, 243, 253, 267, 312, 348, 392, 398, 415, 458, 499, 505, 544, 576, 585
ND6	J	36, 60, 63, 64, 72	26, 59, 74, 117, 132	25, 33, 58, 79, 82, 101, 112, 119, 171
NDUFA1	A	8, 37	19	32, 43
NDUFA2	S	-	44	68
NDUFA3	B	-	-	-
NDUFA5	V	-	-	-
NDUFA6	W	-	[3]	-
NDUFA7	R	-	-	-
NDUFA8	X	-	108	-
NDUFA9	P	286	-	26, 33, 40, 50, 279, 287, 303
NDUFA10	O	107	64, 259	100, 173, 176, 251
NDUFA11	Y	-	-	9, 77, 81, 103
NDUFA12	Q	-	-	-
NDUFA13	Z	[4]	56	-
NDUFAB1	U/T	-	-	-
NDUFB1	F	-	-	-
NDUFB2	J	-	-	-
NDUFB3	K	-	21	[6]
NDUFB4	M	-	-	-
NDUFB5	H	-	-	-
NDUFB6	I	-	-	-
NDUFB7	O	-	-	-
NDUFB8	L	-	-	-
NDUFB9	N	-	63, 96	60, 157
NDUFB10	P	-	-	-
NDUFB11	G	-	92	-
NDUFC1	C	-	-	-
NDUFC2	D	-	-	-
NDUFS4	Q	-	72, 77	[6], 42
NDUFS5	E	-	95	-
NDUFS6	R	87	-	49, 55, 58, 60, 74
NDUFV3	S	-	[22]	-

Table 6.1 Pathogenic variants in complex I subunit genes. Variants were classed by confidence of their assignment as described above, with class 1 the highest confidence class and class 3 the lowest confidence. Variants in brackets are residues not present in the structure.

Subunit	Chain	Class 1	Class 2	Class 3
NDUFV1	F	-	-	[1], [7], 11, 32, 133, 168, 175, 179, 337, 339, 366, 376, 385, 418
NDUFV2	E	102	8	19, 76, 140
NDUFS1	G	-	408	6, 122, 154, 181, 280, 331, 334, 479, 483, 511, 515, 634, 646, 689
NDUFS2	D	319	290	80, 259, 287, 347, 379, 393, 409
NDUFS3	C	-	123	100, 210
NDUFS7	B	-	-	[4], [11], [15], 133, 134, 148, 167
NDUFS8	I	-	-	-
ND1	H	4, 30, 67, 81, 164, 240, 304, 309	1, 10, 11, 29, 31, 39, 43, 64, 69, 73, 87, 89, 113, 144, 147, 187, 229, 236, 248, 263, 275, 277, 289, 307	2, 5, 15, 17, 21, 27, 28, 52, 56, 68, 76, 77, 96, 110, 112, 116, 128, 132, 139, 152, 153, 168, 171, 176, 196, 213, 215, 230, 239, 241, 249, 258, 260, 265, 266, 268, 273, 276, 292, 300, 305, 310, 313, 318
ND2	N	8, 119, 150, 193, 237, 325, 331	13, 31, 43, 57, 64, 76, 80, 88, 89, 100, 122, 152, 164, 202, 203, 265, 270, 278, 284, 288, 323, 342	11, 15, 19, 48, 49, 62, 69, 78, 79, 81, 85, 94, 98, 115, 125, 139, 141, 145, 147, 148, 156, 159, 166, 187, 200, 209, 219, 220, 222, 238, 239, 241, 242, 246, 266, 267, 276, 277, 281, 285, 286, 317, 320, 328, 330, 332, 333, 343, 345, 346
ND3	A	88, 114	9, 10, 29, 45, 49, 60, 96, 103	14, 19, 24, 34, 44, 47, 89, 90, 93, 97, 105
ND4	M	140, 404	47, 50, 54, 86, 89, 101, 109, 110, 131, 138, 149, 165, 230, 299, 313, 340, 390, 402, 423, 424, 442, 445, 459	6, 21, 29, 45, 46, 52, 55, 58, 70, 85, 104, 107, 121, 129, 166, 170, 187, 193, 201, 225, 232, 291, 317, 320, 350, 365, 391, 409, 418, 419, 420, 421, 425, 434, 435, 455
ND4L	K	47	13, 62, 94	21, 59, 71, 79, 80
ND5	L	8, 24, 257, 314, 458, 475, 482, 485, 515, 531, 533, 544, 555	1, 4, 9, 13, 21, 22, 23, 52, 69, 74, 100, 159, 182, 190, 201, 202, 205, 261, 265, 267, 270, 283, 398, 410, 432, 434, 439, 449, 459, 469, 471, 478, 489, 495, 518, 530, 536, 538, 541, 549, 556, 573, 575, 576, 592, 598	2, 14, 18, 27, 30, 34, 38, 39, 40, 45, 53, 59, 63, 65, 67, 70, 76, 88, 93, 99, 109, 113, 127, 132, 141, 147, 149, 162, 169, 172, 177, 196, 206, 209, 211, 236, 271, 273, 290, 304, 317, 323, 331, 377, 378, 381, 400, 403, 413, 420, 423, 424, 440, 441, 442, 447, 456, 461, 472, 473, 476, 477, 491, 492, 493, 500, 508, 509, 513, 517, 524, 525, 546, 564, 565, 569, 571, 577, 579, 596, 602, 603
ND6	J	119, 166,	14, 31, 33, 41, 58, 64, 86, 97, 120, 132, 139, 156, 159, 165, 171	2, 4, 26, 34, 35, 36, 37, 38, 42, 45, 54, 72, 81, 85, 90, 94, 102, 106, 112, 117, 121, 134, 135, 140, 142, 150, 155, 161, 162, 174
NDUFA1	a	-	32	-
NDUFA2	S	-	13	[10], 34, 62, 63, 93
NDUFA3	b	-	20	11, 49
NDUFA5	V	-	-	2
NDUFA6	W	[8]	107	63
NDUFA7	r	47	65	4, 7, 29, 37, 50, 57, 59
NDUFA8	X	-	-	43, 54, 134
NDUFA9	P	259	40	21, 50, 135, 187, 206, 235, 249, 279, 342
NDUFA10	O	-	30	2, 100, 137, 142, 165, 182, 251
NDUFA11	Y	-	103	19, 65, 99, 111, 128
NDUFA12	q	-	-	60
NDUFA13	Z	-	-	9, 24, 41, 69, 80, 87
NDUFAB1	U/T	-	-	-
NDUFB1	f	-	18	33, 38
NDUFB2	j	-	52	10
NDUFB3	k	-	[6]	21, 60
NDUFB4	m	-	-	2, 25, 66
NDUFB5	h	-	-	11, 87, 137
NDUFB6	i	-	-	10, 20, 25, 42, 80, 108
NDUFB7	o	-	62, [117]	[1], 29, 43, 48, 55, 61, 70, 83, 92, 105, 110, 112
NDUFB8	l	-	97, 121	29, 39, 76, 88, 94, 95
NDUFB9	n	145	-	3, 7, 36, 46, 128, 143, 157, 174, 176, 178
NDUFB10	p	-	141	34, 65, 109, 139, 164, [167]
NDUFB11	g	-	-	[13], [22], 55, 56
NDUFC1	c	-	-	41
NDUFC2	d	46	-	35, 59, 119
NDUFS4	Q	-	-	94, 115
NDUFS5	e	-	-	71, 92, 97
NDUFS6	R	-	-	58, 60
NDUFV3	s	-	-	-

Table 6.2 Polymorphic variants in complex I subunit genes. Variants were classed by confidence of their assignment as described above, with class 1 the highest confidence class and class 3 the lowest confidence. Variants in brackets are residues not present in the structure.

	Class 1 polymorphisms	Class 2 polymorphisms	Class 3 polymorphisms	Not observed as polymorphisms	Class total
Class 1 mutations	0	4	9	38	51
Class 2 mutations	3	13	9	61	86
Class 3 mutations	7	24	28	75	134
Not observed as mutations	31	118	362		
Class total	41	159	408		

Table 6.3 Distribution of variants across the six classes and overlap between the lower confidence classes. No overlap between the class 1 variants demonstrates the high level of confidence whereas in other classes there is some degree of overlap between variants, assigned as both pathogenic and polymorphic.

Table 6.3 demonstrates that there was no overlap between the class 1 mutations and the class 1 polymorphisms, confirming that these groups are the most reliably defined. Among the other categories a variable degree of overlap was observed and therefore class 1 variants were distinguished from the remaining dataset in subsequent analyses. Class 1 analyses are presented, which are cleaner but statistically limited by small sample sizes, alongside analyses of the larger, but less well defined, complete dataset. The overlap between Class 2 and 3 mutations and polymorphisms may arise from two effects. Clinical variants may be mis-assigned as pathogenic when a novel variant is identified in a patient with an appropriate biochemical defect but subsequently sequencing a larger control group demonstrates the presence of the same variant in healthy subjects. Additionally, the boundary between mutation and polymorphism is not definitive due to the impact of different biological and metabolic backgrounds: the same mild variant may present differently in individuals (additional factors may give rise to clinical symptoms, or the same variant may remain completely benign).

6.4.Variant distribution across different classes of subunits

Human complex I comprises 45 subunits, 38 are nuclear-encoded (nDNA) and seven are encoded by mtDNA. The 14 core catalytic subunits comprise the mtDNA-encoded proteins and seven nDNA encoded proteins, with the remaining 31 supernumerary subunits likely to have roles in stability, assembly and supercomplex formation. Although the core subunits only comprise 53% of the total residues, 90% of class 1 mutations, and 84% of all mutations, are found these subunits; as expected due to their essential role in complex I function. However, it was surprising to observe a high proportion of polymorphisms also found in the core subunits, 90% of class 1 and 79% of all polymorphisms.

Figure 6.1 shows the distribution of variants across different types of subunits - mitochondrial-encoded, nuclear-encoded core and supernumerary subunits, with raw data in Table 6.4. The percentage of the

residues in each grouping found to be mutations or polymorphisms was compared to the expected segregation of variants. From this analysis it is clear that the mitochondrial-encoded subunits are responsible for the higher polymorphic variation among core subunits, 1.7% of residues are class 1 polymorphisms, which is significantly higher than the 0.6% expected by chance. In contrast, nuclear-encoded core subunits contained a significantly lower proportion than expected, 0.1% - which was equivalent to that of the supernumerary subunits. The distribution of mutations are more similar, 6% of residues in mitochondrially encoded core subunits were found as mutations, along with 4.4% of nuclear encoded subunits, which is higher than in supernumerary subunits, only 1.1%. Additionally, there was a considerable difference found in the type of variants found in the two core subunits groupings, 68% of all variants in the nuclear-encoded core subunits were mutations but only 23% in the mitochondrial-encoded subunits variation and 26% in supernumerary subunits. Overall, this indicates that mtDNA-encoded subunits are more tolerant to variation, which correlates with their lower conservation (38% in mammals) compared to nuclear encoded subunits (78%) and reflects the faster evolution of the mitochondrial genome. However it is difficult to interpret why variants in mtDNA core subunits are less likely to be pathogenic than nuclear encoded core subunits, although one possibility may be their position in the hydrophobic domain. Consequently these subunits have a higher proportion of transmembrane helices (TMHs) and these structures may be less subject to structural disruption than globular subunits, such as the nuclear-encoded core subunits.

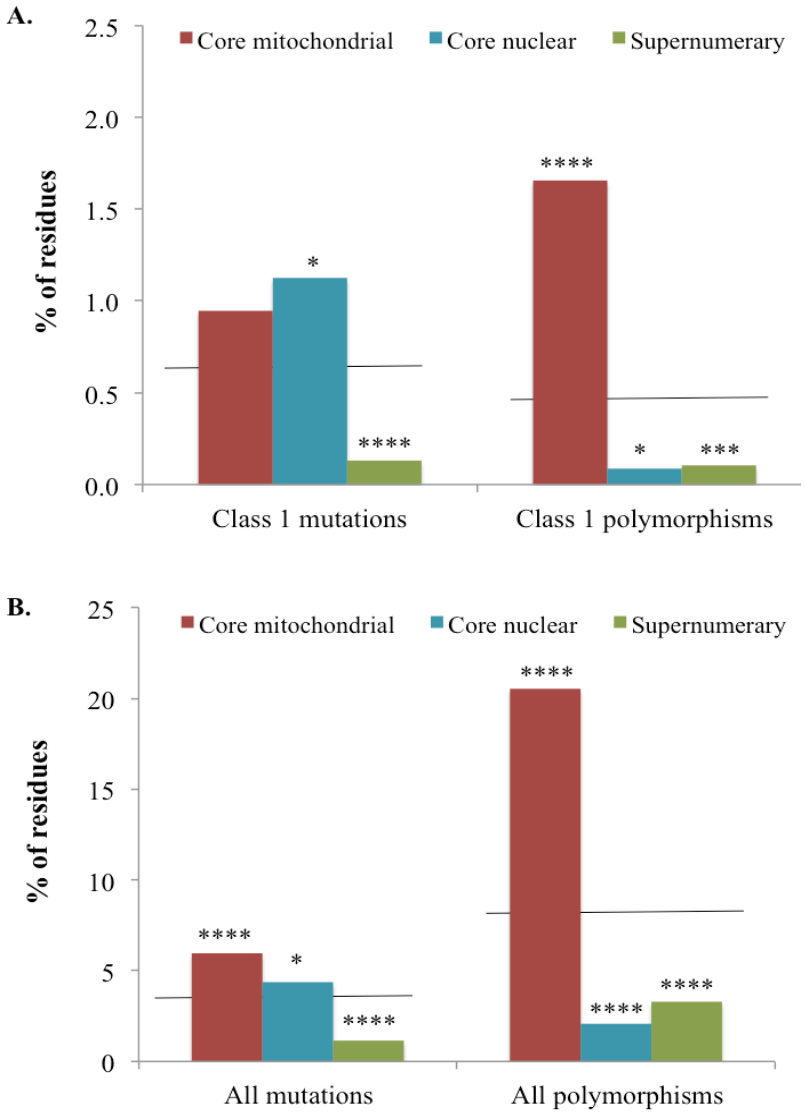


Figure 6.1 The percentage of residues present as mutations and polymorphisms in different groupings of subunits. A. Class 1 mutations and polymorphisms. Data for core mitochondrial subunits is shown in red, nuclear-encoded core subunits in blue and supernumerary subunits in green. Line shows the percentage of residues expected in each grouping if distribution was due to 'chance.' **B. Mutations and polymorphisms in all of classes 1, 2 and 3.** The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences between the groups and that expected by chance was assessed using Fischer's Exact Test (two-tailed) and denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.4.

Subunit	Mutations		Polymorphisms		All residues
	Class 1	Total	Class 1	Total	
mtDNA-encoded	20	126	35	434	2114
nDNA-encoded	26	101	2	48	2311
Supernumerary	5	44	4	126	3828

Table 6.4 Distribution of mutations and polymorphisms across different groupings of subunits. The number of class 1 mutations and polymorphisms found in each type of subunit, along with the total for each variant type. The proportion of all residues in the structure present in each subunit type are also shown, for comparision.

6.5. Primary sequence analysis

Residue conservation was assessed by aligning subunit sequences from twenty mammalian species, chosen for their diversity and availability of sequences. Table 6.5 shows the list of species and where necessary, predicted sequences or sequences from closely related species were used, leading to a total of 867 sequences included in the analyses. Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers *et al*., 2011) was used to align the sequences from each subunit and the conservation of each residue classified by four categories: fully conserved (identical in all species), strongly conserved (all within a strong conservation group of similar residues), weakly conserved (all within a weak group of similar residues) and non-conserved. These levels of conservation were calculated using a point accepted mutations (PAM) matrix. PAM matrices are constructed by analysing substitution frequencies in evolutionary related proteins - to determine the probability that a given residue will be replaced by another particular residue, by chance, after a set evolutionary interval. PAM units reflect evolutionary distance, so that for the PAM250 matrix only an average of one in five residues remain unchanged. Using the Gonnet PAM250 matrix, Clustal Omega scored each residue position based on its mutation probability, with a score of 1 reflecting full conservation, 1 – 0.5 strong conservation, 0.5 – 0 weak conservation, and 0 not conserved. The groups of strong and weak conservation are given in Table 6.6.

For analysis of the severity of amino acid change each variant was considered individually, with residue positions demonstrating more than one change separated and reclassified (Appendix 7.9D). The specific amino acid change was then assigned as severe, medium or mild. Severe represented a non-conserved change, medium was a change that remained within a weakly conserved residue grouping and a mild change remained within a strongly conserved residue grouping, according to Table 6.6.

Species	Common name	Notes
Bos taurus	Cow	All subunit sequences fully annotated
Gorilla gorilla gorilla	Western lowland gorilla	All subunit sequences fully annotated
Homo sapiens	Human	All subunit sequences fully annotated
Macaca mulatta	Rhesus monkey	All subunit sequences fully annotated
Mus musculus	Mouse	All subunit sequences fully annotated
Pan troglodytes	Chimpanzee	All subunit sequences fully annotated
Rattus norvegicus	Rat	All subunit sequences fully annotated
Ailuropoda melanoleuca	Giant panda	Predicted sequence used for NDUFV3, NDUFS1, NDUFS4, NDUFA12, NDUFA13, NDUFAB1, NDUFB2, NDUFB5
Callithrix jacchus	White-tufted-ear marmoset	Predicted sequence used for NDUFA3, NDUFA6, NDUFA7, NDUFA10, NDUFA11, NDUFB5, NDUFC2, NDUFV3
Canis familiaris	Dog	Predicted sequence used for NDUFS5, NDUFA1, NDUFA10, NDUFB1, NDUFB5, NDUFB8
Cavia porcellus	Guinea pig	Predicted sequence for NDUFV3, NDUFS7, NDUFA1, NDUFA7, NDUFA12, NDUFB1, NDUFB4, NDUFC2
Equus caballus	Horse	Predicted sequence used for NDUFA7, NDUFA11, NDUFA12, NDUFA13 (from *Equus przewalskii)*, NDUFB4, NDUFB6, NDUFV3 (from *Equus asinus)*
Felis catus	Cat	Predicted sequence used for NDUFS6, NDUFA3, NDUFA12, NDUFA13, NDUFB1 No sequence for NDUFV3 Sequences for NDUFC2, NDUFS2 incorrect (discounted)
Ictidomys tricemlieantus	Thirteen-striped ground squirrel	Predicted sequence used for NDUFV3, NDUFA2, NDUFA8, NDUFA9, NDUFA11, NDUFA12, NDUFA13, NDUFB1, NDUFB4, NDUFB5, NDUFC2 Sequence for NDUFA12 incorrect (discounted)
Monodelphis domestica	Opossum	Predicted sequence used for NDUFS8, NDUFA12, NDUFB2, NDUFC2 No sequence for NDUFV3 Sequence for NDUFS8 incorrect (discounted)
Myotis lucifugus	Little brown bat	Predicted sequence used for NDUFV3, NDUFS1 (from *Myotis brandtii*) NDUFS4, NDUFA8, NDUFA9, NDUFA10, NDUFA11, NDUFA12, NDUFB2, NDUFC2 NDUFA5 from *Myotis brandtii* No sequence for NDUFA3
Ovis aries	Sheep	Predicted sequence for NDUFV3, NDUFS2, NDUFS6, NDUFS8 (from *Ovis musimon*), NDUFA5, NDUFA13, NDUFC1
Pongo pygmaeus	Bornean orangutan	*Pongo abelii* sequences used for NDUFA1, NDUFA3, NDUFB9
Sarcophilus harrisii	Tasmanian devil	Predicted sequence used for NDUFV3, NDUFS1, NDUFS4, NDUFS6, NDUFA2, NDUFA8, NDUFA12, NDUFA13, NDUFB1, NDUFB5, NDUFB7, NDUFC1 No sequence for NDUFS3, NDUFB6, NDUFB9 Sequence for NDUFB7, NDUFC2 incorrect (discounted)
Sus scrofa	Pig	Predicted sequence used for NDUFS1, NDUFS8, NDUFA12, NDUFB7, NDUFB9 No sequence for NDUFV3

Table 6.5 Species used for sequence alignment by Clustal Omega to determine conservation of complex I subunits. For subunits without annotated sequence in a given species predicted sequences or sequences of closely related species were used where available.

Strong conservation groups	Weak conservation groups
STA	ATV
NEQK	SAG
NHQK	STNK
NDEQ	STPA
QHRK	SGND
MILV	SNDEQK
MILF	NDEQHK
HY	NEQHRK
FYW	FVLIM
	HFY

Table 6.6 Gonnet PAM250 grouping of residues for strongly and weakly conserved conservation assignment. These groupings were also used for the severity of amino acid change – strongly conserved were mild changes and weakly conserved were medium changes, with non-conserved change denoted as severe.

6.5.1. Conservation

Figure 6.2 demonstrates the distribution of variants across the four classes of conservation, with raw data shown in Table 6.7. There is a clear and significant trend in the proportion of variants that are fully conserved and not conserved across the classes. Class 1 mutations are in residues that are significantly more conserved than class 1 polymorphisms, 88% compared to 5%, whereas a significant proportion of class 1 polymorphisms are not conserved (71%) compared to class 1 mutations (4%). Additionally, the data shows that for fully conserved residues there is segregation towards pathogenic mutations and against benign polymorphisms. 55% of all the residues in the structure are in fully conserved residues; therefore there are more than 30% more class 1 mutations and 50% fewer class 1 polymorphisms that are fully conserved than expected by chance. The opposite trend was observed for non-conserved residues; this conservation level makes up less than 25% of the total residues, but the proportion is considerably higher for class 1 polymorphisms and far lower for class 1 mutations. The intermediate classes of strongly and weakly conserved showed less pronounced trends and less difference between class 1 variants, therefore the predictive power of conservation levels can be simplified to considering whether the residue is fully conserved or not. This analysis shows that variants can be assessed for likely pathogenicity from their conservation level, not only is there a significant difference between the two highest confidence classes but a trend was also observed across all six variant classes.

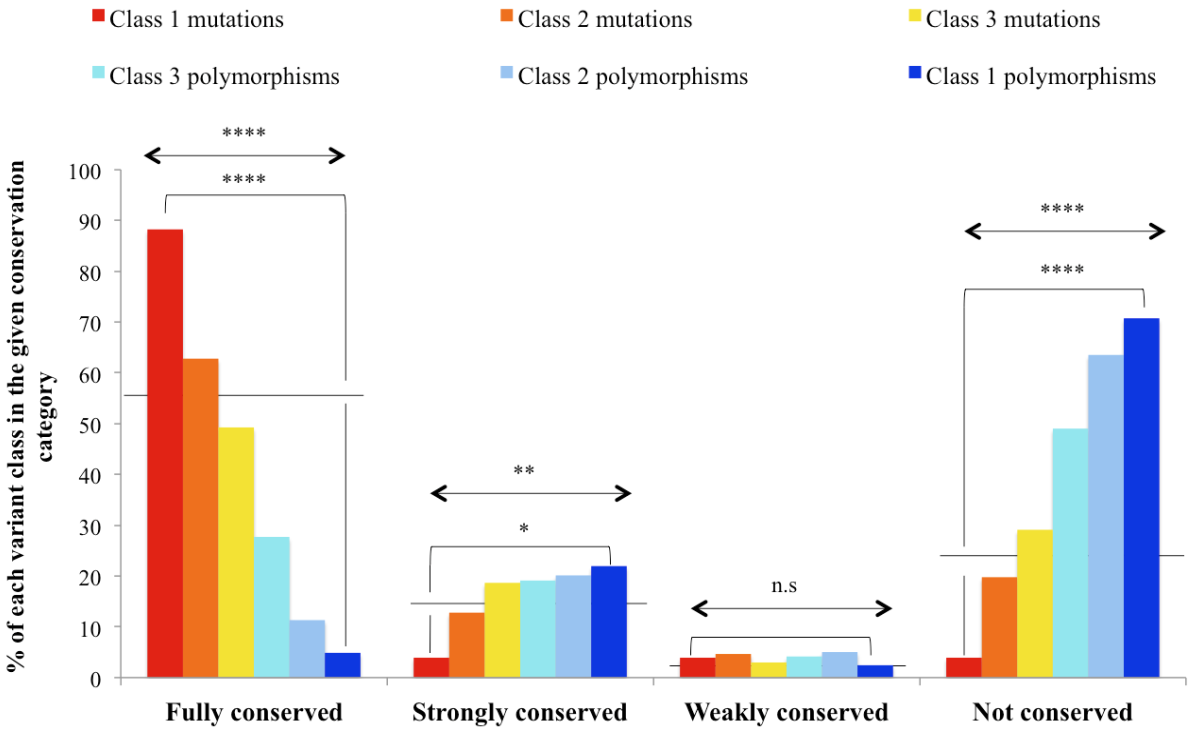


Figure 6.2 Trends in residue conservation across the classes of mutations and polymorphisms. Conservation categories were determined by Clustal Omega as described in above. The proportion of variants that were fully, strongly, weakly or not conserved are shown in the order class 1, 2, 3 mutations and class 3, 2, 1 polymorphisms for trend analysis. Conservation levels across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.7.

	Mutations			Polymorphisms			All residues
	Class 1	Class 2	Class 3	Class 3	Class 2	Class 1	
Fully conserved	45	54	66	113	18	2	4505
Strongly conserved	2	11	25	78	32	9	1414
Weakly conserved	2	4	4	17	8	1	358
Not conserved	2	17	39	200	101	29	1976

Table 6.7 Distribution of mutations and polymorphisms across different conservation categories. The number of variants from each class found in conservation category. The proportion of all residues in the structure present in each conservation level is also shown, for comparision.

Figure 6.3 shows the conservation level of residues mapped onto the human complex I model (generated as described in section 6.6) and demonstrates the hydrophobic domain is less conserved than the hydrophilic arm. This correlates well with the observation that ~75% of class 1 polymorphisms are non-conserved because, as highlighted in Figure 6.1, polymorphisms are significantly more likely in the mitochondrial core subunits of the membrane arm, which demonstrates low conservation of residues.

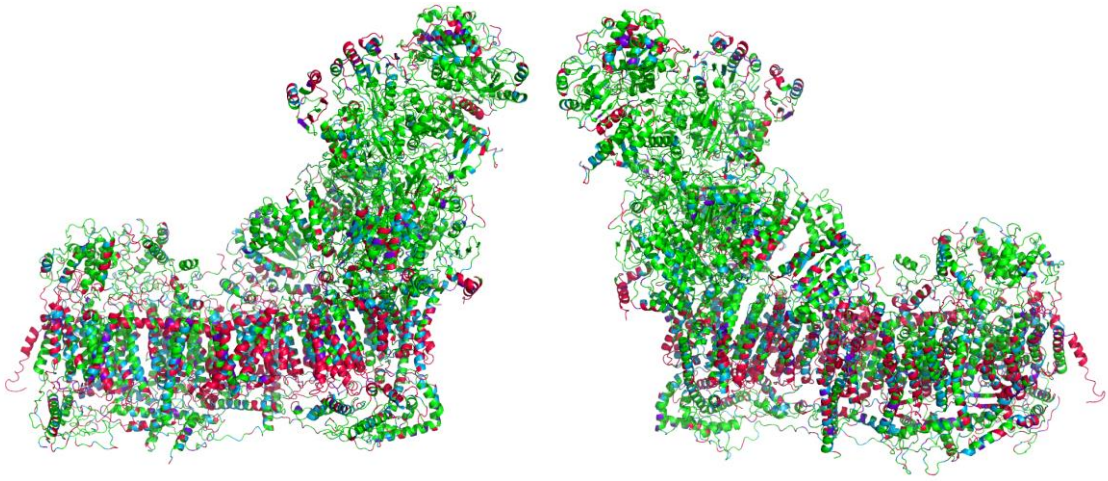


Figure 6.3 Structure of human complex I coloured by conservation. The conservation of all residues in the structure was determined by alignment of subunit sequences for 20 mammalian species (Table 6.5). Green denotes fully conserved residues, blue are strongly conserved, purple are weakly conserved and red are non-conserved residues. Both side views of the model are shown. The human model of complex I was constructed using mouse complex I cryo-EM structures (section 6.6).

6.5.2. Severity of amino acid change

Another important, clinically relevant consideration is the type of amino acid change. The severity of this change will have implications for the properties of the residue, directly affecting its function – therefore not all changes to a residue will have the same impact. For this analysis any residue positions with multiple changes were considered separately, so the classification of some residues was altered, along with the total number of residues in each variant class. Figure 6.4 shows the proportion of residues from each variant class in the different amino acid change categories, denoted severe for non-conserved changes, medium from weakly conserved changes and mild for strongly conserved changes. The data demonstrates that a variant changing to a very different residue is more likely to result in a pathogenic mutation; there is a significant trend across the six classes and significantly more class 1 mutations are found with a severe change compared to class 1 polymorphisms. A medium change in residue property did not show a clear trend and suggests that this type of change may result in either mutations or polymorphisms. However polymorphisms are more likely to arise from a mild change, causing little difference in the property of the amino acid, compared to a mutation. There was a significant increasing trend across the variant classes (mutations to polymorphisms), with less than 50% of class 1 mutations

accounted for by mild mutations whereas class 1 polymorphisms comprised a large proportion, 73%. Therefore the predictive power of amino acid change could be considered as binary - whether the residue change is severe or not.

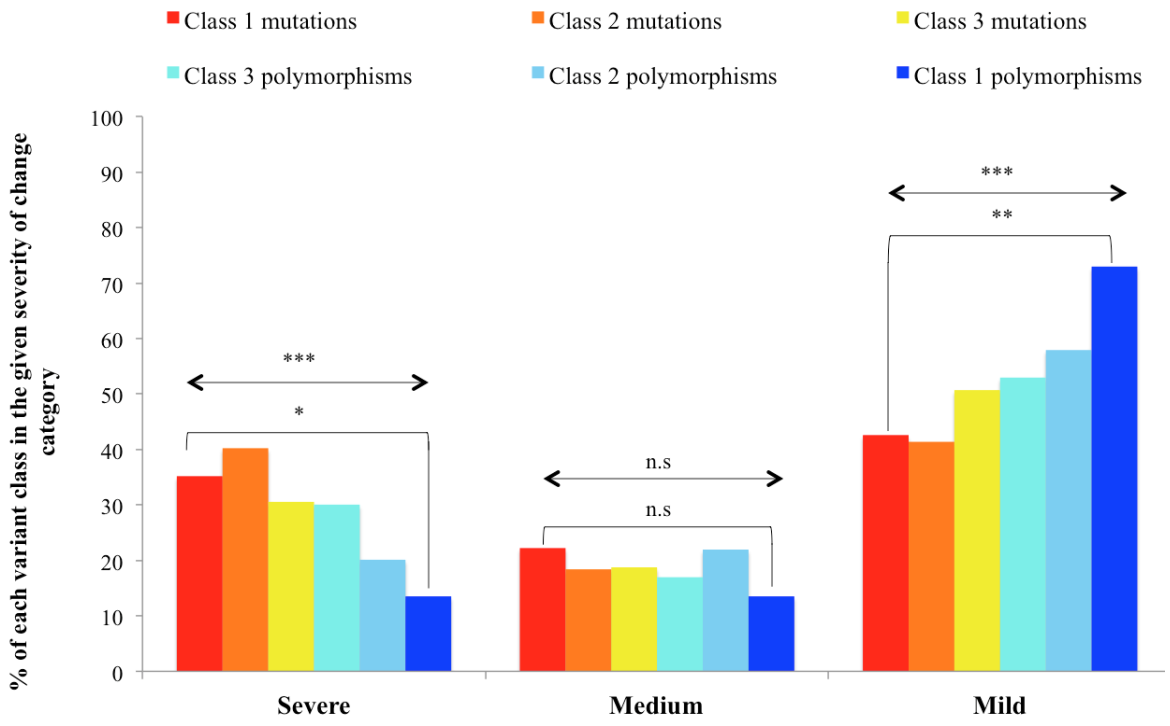


Figure 6.4 Trends in residue change severity across the classes of mutations and polymorphisms. Severity of change categories were determined by Clustal Omega as described in above. The proportion of variants with severe, medium and mild changes are shown in the order class 1, 2, 3 mutations and class 3, 2, 1 polymorphisms for trend analysis. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.8.

	Mutations			Polymorphisms		
	Class 1	**Class 2**	**Class 3**	**Class 3**	**Class 2**	**Class 1**
Severe	19	34	44	147	33	5
Medium	12	16	27	83	36	5
Mild	23	37	73	259	95	27
Total	54	87	144	489	164	37

Table 6.8 Distribution of mutations and polymorphisms with residue changes of different severity. Severe changes are non-conserved, medium changes are weakly conserved and mild changes are strongly conserved as determined by residue groupings from Clustal Omega, Table 6.6.

6.6. The high-resolution model of human complex I provides more information about residues in the structure

Model building was carried out by Dan Grba from the MRC MBU, University of Cambridge. The structure of mouse complex I in the active state (RCSB protein data bank (PDB) code: 6G2J; Agip *et al*., 2018), the highest-resolution structure of mammalian complex I currently available, was used as the basis for the structural model of human complex I. First, sequence changes (additions, deletions and mutations) were performed in Coot (version 0.8.9.1-pre, Emsley *et al*., 2010), using the curated sequences of the human complex I subunits in the UniProt/Swiss-Prot database and mouse-human sequence alignments from LALIGN (https://www.ebi.ac.uk/Tools/psa/lalign/). The sequence identity between the mouse and human complexes is 81%. The humanised model was then refined against the cryo-EM electron density for the mouse enzyme (EMD-4345, Agip *et al*., 2018) by real-space refinement in Phenix (version 1.12-2829, Adams *et al*., 2010) and geometrical restraints were used for the iron-sulphur clusters. Phospholipids were not included. It was then placed into the lower-resolution complex I portion of the human supercomplex cryo-EM density map (EMD-6775 and EMD-6773, Guo *et al*., 2017) using UCSF Chimera (version 1.11.2, Petteresen *et al*., 2004). Using Phenix, subunit positions were adjusted by rigid body fitting and then real-space refinement was performed. Minimal side-chain movements were observed as a result, particularly in poorly resolved regions. In a small number of regions the densities from the supercomplex map or a subsection of it (Guo *et al*., 2017) were better defined than the mouse density, allowing a small number of additional structures to be built using Coot. The final model was produced by manual adjustment and real-space refinement and validated using MolProbity (http://molprobity.biochem.duke.edu, Williams *et al*., 2018) and EMRinger (http://fraserlab.com/2015/02/18/EMringer/, Barad *et al*., 2015). The overall root-mean-square deviation between our model and the model presented by Guo and co-workers is very low at 0.8 Å, showing high general similarity, with the weakest agreement for NDUFV1 and NDUFV2 (1.2 Å, two subunits of relatively low resolution in the supercomplex density) and the best agreement (0.5 - 0.6 Å) in the heel and centre of the membrane domain. The human model generated from the structure of the active state of mouse complex I was used for the analyses described because it originates from a higher-resolution analysis than the model of Guo *et al.* and describes an enzyme in a known functional state; additionally we have detailed knowledge and understanding of it. The model is summarised in Table 6.9.

Subunit	Chain	Total residues	Modeled residues	% Modeled	% Identity	Number of residues cleaved from N-terminus	Notes
NDUFV1	F	444	10-444	98.0	97.3	20	FMN, 4Fe4S
NDUFV2	E	217	1-217	100	96.3	32	2Fe2S
NDUFS1	G	704	6-693	97.7	94.7	23	2Fe2S, 2 x 4Fe4S
NDUFS2	D	430	1-430	100	93.7	33	Dimethyl-Arg85
NDUFS3	C	228	8-215	91.2	91.2	36	
NDUFS7	B	175	19-175	89.7	92.0	38	4Fe4S
NDUFS8	I	176	1-176	100	94.9	34	2 x 4Fe4S
ND1	H	318	1-318	100	77.0	0	N-formyl
ND2	N	347	1-347	100	57.1	0	N-formyl
ND3	A	115	1-115	100	66.1	0	N-formyl
ND4	M	459	1-459	100	66.9	0	N-formyl
ND4L	K	98	1-98	100	66.3	0	N-formyl
ND5	L	603	1-603	100	63.8	0	N-formyl
ND6	J	174	1-174	100	52.3	0	N-formyl
NDUFV3	s	74	34-74	55.4	68.9	34	
NDUFS4	Q	133	9-133	94.0	94.0	42	
NDUFS5	e	105	1-105	100	74.3	1	2 x Cys-Cys
NDUFS6	R	96	1-96	100	88.5	28	Zn^{2+}
NDUFA1	a	70	1-70	100	82.9	0	
NDUFA2	S	98	13-95	84.7	85.7	1	
NDUFA3	b	83	1-83	100	80.7	1	N-acetyl
NDUFA5	V	115	1-115	100	82.6	1	
NDUFA6	W	127	14-127	89.8	74.0	27	
NDUFA7	r	112	1-76, 90-122	88.4	91.1	1	N-acetyl
NDUFA8	X	171	1-171	100	86.5	1	4 x Cys-Cys
NDUFA9	P	342	1-342	100	79.5	35	NADPH
NDUFA10	O	320	1-320	100	77.5	35	ADP
NDUFA11	Y	140	1-140	100	65.0	1	Cys-Cys
NDUFA12	q	145	1-145	100	89.0	0	N-acetyl
NDUFA13	Z	143	5-143	97.2	83.2	1	
NDUFAB1α	T	88	8-82	85.2	96.6	68	Acyl chain
NDUFAB1β	U	88	1-88	100	96.6	68	Acyl chain
NDUFB1	f	57	1-57	100	75.4	1	
NDUFB2	j	72	1-71	98.6	87.5	33	
NDUFB3	k	97	14-92	81.4	82.5	1	
NDUFB4	m	128	1-128	100	75.0	1	
NDUFB5	h	143	1-143	100	77.6	46	
NDUFB6	i	127	1-127	100	70.1	1	N-acetyl
NDUFB7	o	136	12-116	77.2	82.4	1	2 x Cys-Cys
NDUFB8	l	157	1-154	98.1	84.7	29	
NDUFB9	n	178	1-178	100	86.5	1	
NDUFB10	p	171	1-165	96.5	80.7	1	2 x Cys-Cys
NDUFB11	g	124	23-124	82.3	78.2	29	
NDUFC1	c	49	1-49	100	75.5	27	
NDUFC2	d	119	1-119	100	68.9	0	

Table 6.9 Summary of structure models for subunits in human complex I. Some subunits include an N-terminal mitochondrial targeting sequence (MTS) or initiator methionine that is cleaved; the number of residues removed is given and indicates the change in amino acid numbering in the mature protein. The percentage identity denotes the sequence similarity between the mouse and human subunit sequences.

6.6.1. Secondary structure

The four secondary structure elements considered are: alpha helices, beta sheets, transmembrane helices and loops. The secondary structure elements of proteins have different properties and may be associated with different functions. Helices that span membranes are termed transmembrane helices. Loops between other elements are highly flexible and may undergo large conformational changes. Additionally, different secondary structure elements often associate to form a distinct fold, with combinations of beta sheet and alpha helices used to form the nucleotide binding Rossmann fold, and TIM barrels are common in the catalytic domains of enzymes (Petsko & Ringe, 2003). Secondary structure may be an important predictor of pathogenicity because some elements may be more tolerant to variation than others. The secondary structure of each residue was defined in the header of the PDB coordinate file, based on the secondary structure of the equivalent residue in the structure of mouse complex I (PBD code: 6G2J; Agip *et al*., 2018).

Figure 6.5 shows the distribution of variants across the four secondary structure categories and comparison with the proportion of each in the whole structure, raw data in Table 6.10. Across complex I loops are the dominant category, comprising almost 50% of the structure, whereas helices and TMHs each account for just over 20%. Only a small population of complex I residues are found in beta sheet structures, just 6%. The six variant classes also show the lowest proportion of residues present in beta sheets and although a significant trend was calculated across the classes, with a higher proportion of mutations, the small size of this category implies that overall variants do not segregate into beta sheet elements; additionally there was no significant difference between the proportion of class 1 mutations and polymorphisms in this secondary structure. Helixes in complex I contain a low proportion of variation overall, with all classes (except class 2 mutations) showing a considerably smaller percentage than the expected 24%. Although a significant trend was calculated across the classes this was only due to the proportion found in class 2 mutations, which was similar to that expected by chance. If this value was removed the analysis was non-significant. Again the difference between class 1 mutations and polymorphisms was also non-significant, indicating that helices generally demonstrated less variation and there was no difference between likelihood of a pathogenic or polymorphic outcome. Although loops are the most prevalent secondary structure these also demonstrated lower than expected levels of variation and there was no significant trend across the classes. Therefore, TMHs appeared to be the only secondary structure showing a difference in variant distribution, with a highly significant trend across the classes (increasing toward polymorphisms). However, even for TMHs, there was no significant difference between class 1 mutations and polymorphisms, but interestingly all classes showed a higher proportion of residues as TMHs than expected, more than 22% - the percentage across the whole structure. Results for TMH elements may be convoluted by the high proportion of TMHs found in the core mtDNA-encoded subunits, therefore separate analysis was carried out.

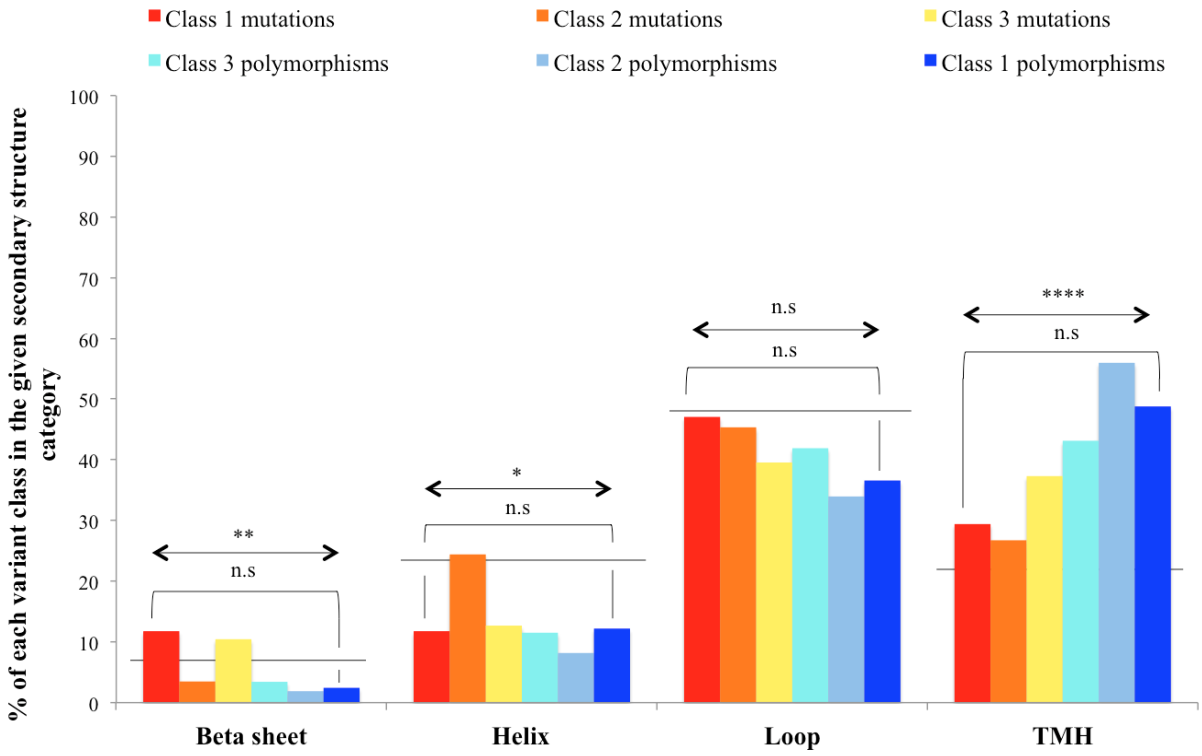


Figure 6.5 Trends in residue secondary structure across the classes of mutations and polymorphisms. The proportions of different secondary structure elements across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.10.

	Mutations			Polymorphisms			All residues
	Class 1	Class 2	Class 3	Class 3	Class 2	Class 1	
Beta sheet	6	3	14	14	3	1	457
Helix	6	21	17	47	13	5	1968
Loop	24	39	53	171	54	15	4003
TMH	15	23	50	176	89	20	1825

Table 6.10 Distribution of mutations and polymorphisms across different secondary structure elements. The number of variants from each class found in the four secondary structure elements. The proportion of all residues in the structure present in each category is also shown for comparision.

Figure 6.6 (and raw data in Table 6.11) shows that the proportion of each variant class found in TMHs of mtDNA-encoded subunits is similar and reflects the percentage expected by chance, 63%. There was no significant trend across the classes or difference between class 1 mutations and polymorphisms. For the nuclear encoded subunits there was a trend across the classes, however there was no significant difference between class 1 mutations and polymorphisms. Although there are fewer mutations than expected compared to the proportion of residues in TMHs across the nuclear subunits, 8%, the difference was not significant – nor was the slight increase in polymorphisms compared to random distribution. Therefore Figure 6.6 show that the distribution of variants in TMHs observed in Figure 6.5 was due to the high proportion of TMH secondary structure in the mtDNA-encoded core subunits, 63% compared to 8% in nuclear-encoded subunits; when separated by genome there was no clear or significant correlation between variant type and TMH elements. This demonstrates that secondary structure does not constitute an independent predictive feature for the likelihood of pathogenic mutations or polymorphisms in complex I.

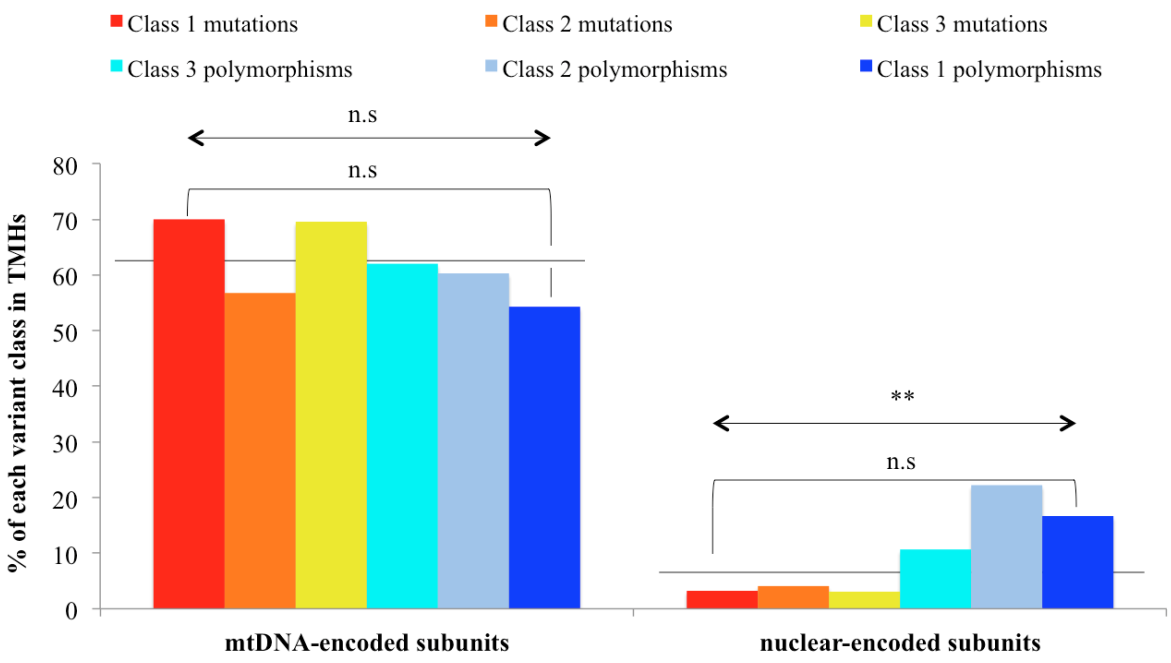


Figure 6.6 Distribution of mutations and polymorphisms across TMHs in mtDNA-encoded and nuclear-encoded subunits. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.11.

	Mutations			Polymorphisms			All residues
	Class 1	**Class 2**	**Class 3**	**Class 3**	**Class 2**	**Class 1**	
mtDNA-encoded	14 (20)	21 (37)	48 (69)	160 (258)	85 (141)	19 (35)	1324 (2114)
nuclear DNA-encoded	1 (31)	2 (49)	2 (65)	16 (150)	4 (18)	1 (6)	501 (6139)

Table 6.11 Distribution of all mutations and polymorphisms across TMHs in subunit groupings. Totals for mutations, polymorphisms and residues in mtDNA and nuclear DNA subunits are given in brackets.

6.6.2. Residue environment

The tertiary structure of a protein is directly linked to its function, as it requires a precise structure to be maintained. Residues are found either on the solvent exposed surface or within the structure, contacting other residues, where they may form important interactions. For multi-subunit proteins like complex I, the boundaries between different subunits may also be important regions that require structural integrity either for function or stability of the complex. To assess variant distribution across different environments, residues within the human model structure were grouped into three categories – surface exposed, buried or at subunit interfaces. The fractional surface areas of each residue in contact with the solvent were calculated using the get_sasa_relative command in PyMol (default probe radius of 1.4 Å). The values are provided as a percentage and calculated as the relative solvent accessible surface area (SASA) compared to the area of the fully exposed residue (with only its two neighbouring residues present). For a residue to be surface exposed it has more than 25% of its surface area accessible to the solvent. Residues with less than 25% of their surface area exposed were non-surface residues. To determine whether these amino acids are at interfaces or completely buried two solvent accessible values were compared: the surface exposed area when the subunit is within the whole enzyme and the area that is solvent accessible when the subunit is in isolation. From these the percentage of the surface area in contact with other subunits was calculated as the difference in values as the relative SASA percentage at a subunit interface is only exposed in the isolated subunit. Thus for non-surface residues, if the percentage exposed only in the isolated subunit was >25% then the residue was deemed to be at an interface, otherwise it was assigned as completely buried in the structure.

Figure 6.7 demonstrates the distribution of variants in the six classes according to their environment, with raw data in Table 6.12. It is clear that variants found in buried and interface environments are more likely to be mutations, whereas those on the surface are more likely to be polymorphic. There is a significant trend in residue environment across the variant classes for buried and surface environments, although the difference between the highest confidence classes was only significant in the surface grouping, with a higher proportion of polymorphisms. Interestingly when compared to the distribution expected by chance, given by the distribution across the whole structure, polymorphisms have a similar

observed distribution, suggesting there is no selective pressure on polymorphisms due to residue environment – 39% of residues in the whole structure are buried, 23% at interfaces and 38% found on the surface – all classes of polymorphisms differ by less than 5% across all residue environments. In contrast, the proportion of class 1 mutations found buried or at interfaces is higher than expected and surface exposed mutations lower than expected, indicating a selective pressure towards variation being pathogenic for non-surface residues.

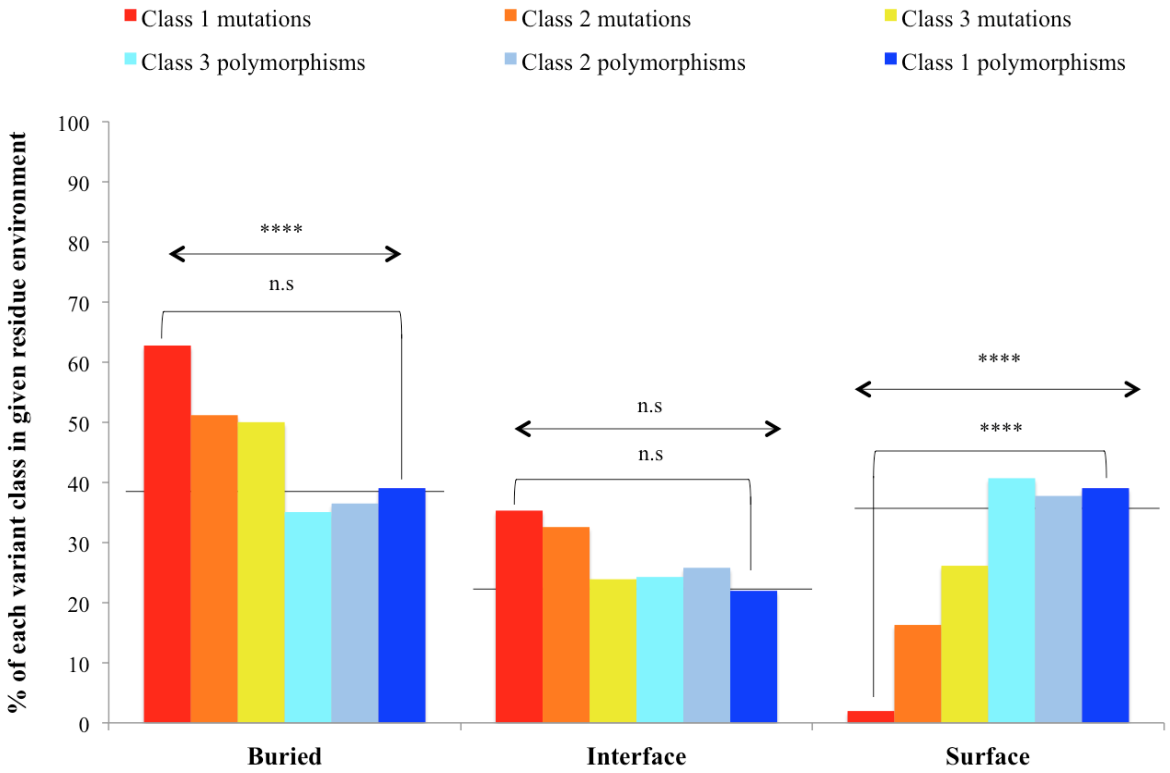


Figure 6.7 Trends in residue environment across the classes of mutations and polymorphisms. The proportions of residues in different environments across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.9.

	Mutations			Polymorphisms			All residues
	Class 1	Class 2	Class 3	Class 3	Class 2	Class 1	
Buried	32	44	67	143	58	16	3240
Interface	18	28	32	99	41	9	1904
Surface	1	14	35	166	60	16	3109

Table 6.12 Distribution of mutations and polymorphisms across different residue environments. Buried and interface residue data was combined for the non-surface grouping in Figure 6.7.

Additionally there appears to be no difference in the segregation of variants between buried and interface environments. Thus these two groupings can be considered together and this feature can be used to predict pathogenicity by simply identifying whether the residue is surface exposed or not, Figure 6.8.

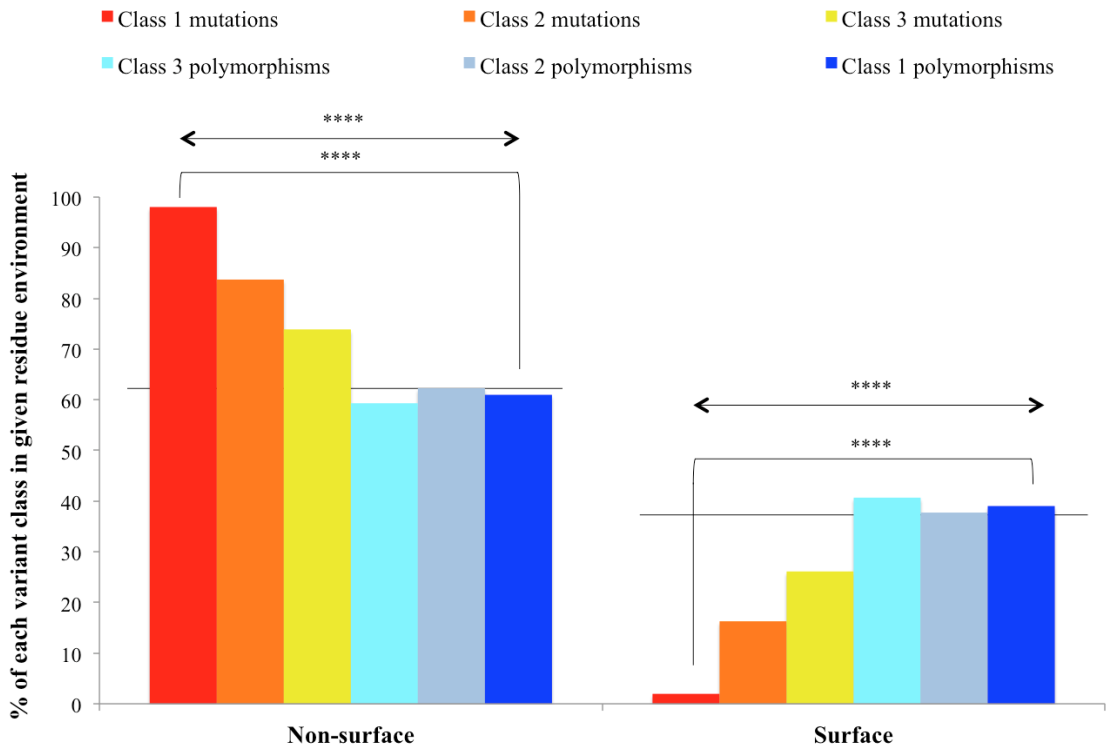


Figure 6.8 Trends in surface and non-surface residue exposure across the classes of mutations and polymorphisms. The proportions of residues in the two environments across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Combining the buried and interface groupings in Table 6.6 gives the data for the non-surface environment.

6.6.3. Salt bridge formation

Another aspect to be considered from structural information is the presence of salt bridges. These are the strongest type of non-covalent interactions and are an electrostatic interaction between two oppositely charged residues (Kumar & Nussiov, 2002). Therefore the types of amino acid were first analysed in each of the variant classes and across the whole structure. Amino acids were grouped by type: aromatic (Phe, Trp, Tyr), negatively charged (Asp, Glu), non-polar (Ala, Gly, Ile, Leu, Met Val), polar non-charged (Asn, Cys, Gln, Pro, Ser, Thr) and positively charged (Arg, His, Lys). Salt bridges were analysed using the ESBRI server (Constantini *et al*., 2008).

Figure 6.9A demonstrates that, across all the variant classes, non-polar residues made up the highest proportion, which reflects the whole structure distribution of 39%. There was no significant trend across the variant classes for aromatic and non-polar residues, although there was a significant trend in polar non-charged residues, which increased from mutations to polymorphisms, despite the difference between class 1 mutations and polymorphisms being non-significant. In contrast, the distribution of charged residues demonstrated a clear trend; no class 1 polymorphisms were negatively charged residues and only one was a positively charged residue, whereas class 1 mutations showed a significantly larger proportion – 22% and 14% respectively. The substantial trends across the variant classes for these two amino acid types and between the class 1 mutations and polymorphisms indicated that, when mutated, charged residues were likely to be pathogenic. To rationalise this observation further the salt bridge forming residues were analysed, 6% of the whole structure comprises charged residues found in a salt bridge pair. The proportion was considerably increased to over 15% for the class 1 mutations, whereas there are no class 1 polymorphisms involved in salt bridges. A strong significant trend was also exhibited across the six classes. Therefore, there is a selection for variation in salt bridge forming residues causing detrimental mutations, Figure 6.9B. Therefore, salt bridges in complex I represent an important structural feature and their disruption is a predictive factor for the outcome of a variant; they may increase structural stability, or be important in bringing functionally important residues together.

	Mutations			Polymorphisms			All residues
	Class 1	**Class 2**	**Class 3**	**Class 3**	**Class 2**	**Class 1**	
Aromatic	1	11	11	20	14	4	875
Negatively charged	7	12	6	9	1	0	884
Non-polar	18	31	67	184	75	20	3239
Polar non-charged	14	16	29	132	63	16	2147
Positively charged	11	16	21	63	6	1	1108
Salt bridge forming	8	10	6	19	1	0	474

Table 6.13 Distribution of mutations and polymorphisms across different amino acid types and the number forming salt bridges. The number of variants from each class found in the different amino acid types. The proportion of all residues in the structure present in each amino acid category is also shown, for comparision.

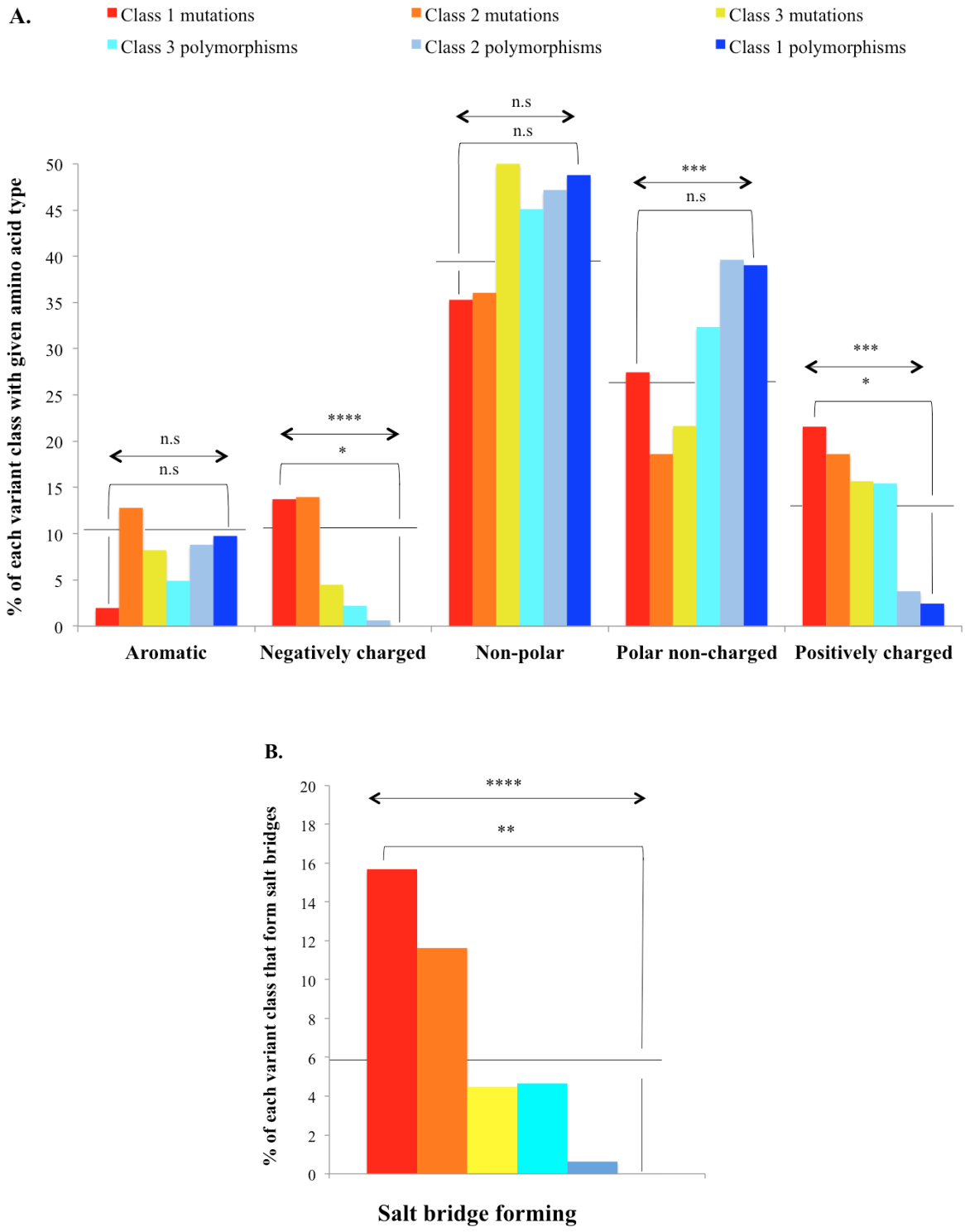


Figure 6.9 Trends in amino acid type across the classes of mutations and polymorphisms. The proportions of residues with different properties across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.10.

6.6.4. Variant distribution across the structure of complex I

Figure 6.10 displays the distribution of variants in the structure of complex I and reflects the observation that mutations and polymorphisms show an unequal spread across the enzyme. In addition, it is clear that, even within groups of subunits, there is no even distribution. Most notably the mtDNA subunits contain a high proportion of class 1 mutations in four of the seven subunits: ND1, ND3, ND4L and ND6 contain 17/20 of the class 1 mutations in the core membrane subunits yet only represent one third of the residues in this group of subunits. Additionally, from visual inspection of Figure 6.10 there are clusters of mutations in two functionally important regions – the N-domain at the top of the hydrophilic arm, where the redox factor FMN is bound, and Q-domain near the base of the hydrophilic arm where it contacts the membrane arm, which contains the proposed ubiquinone-binding site. In contrast, these regions appear to lack polymorphisms and could indicate that variant distribution is influenced by proximity to certain functional domains of complex I. Therefore, to explore this further key features of complex I were identified and the distance from variant residues assessed.

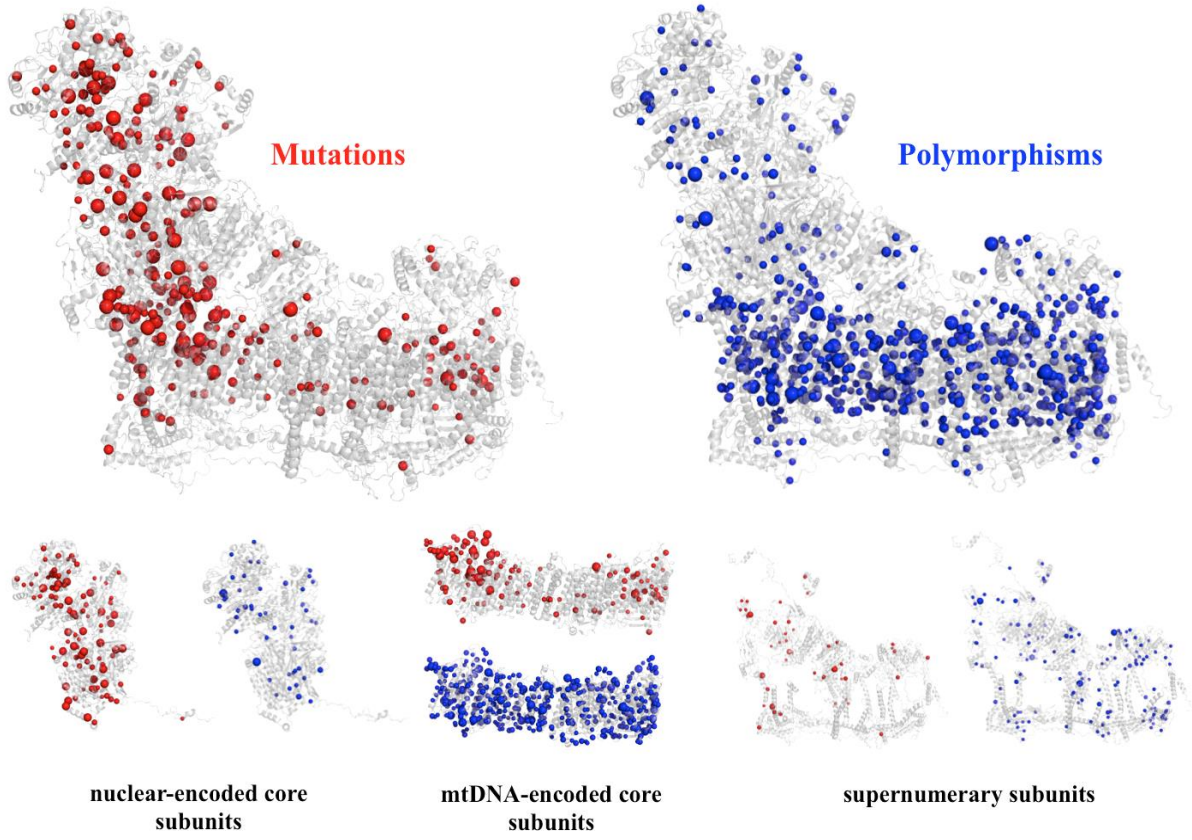


Figure 6.10 Distribution of mutations and polymorphisms in the complex I structure. Mutations are in red and polymorphisms in blue. Class 1 variants are shown by large spheres, class 2 variants by intermediate spheres and class 3 variants by small spheres.

6.6.5. Distance from key features in complex I

Due to its large structure complex I contains a number of functionally and structurally relevant elements. By reviewing the literature and published structures, 16 groups of features were identified (Table 6.14) and the distance of each feature from every residue in the structure was measured. This was achieved using coordinates for the key features: for those made up of residues the alpha-carbon and atoms of the side chain were selected, and for bound molecules coordinates for all atoms were used – details given in Table 6.14. Then the distance of the nearest coordinate from the alpha-carbon of each residue in the structure was found and attributed to the corresponding key feature, to determine the closest feature to each residue. For analyses the features were initially grouped as shown in Table 6.14 into functionally and structurally related elements.

Feature class	**Atom coordinates considered**
Redox cofactors	FMN (*all atoms*), 2Fe2S and 4Fe4S iron sulphur clusters (*all atoms*)
Q-binding site	Ligands for ubiquinone head group - NDUFS2 His59, Tyr108 and Asp160 (*C-α and all side chain atoms*) and 49 midline coordinates for isoprenoid binding channel determined by CAVER
ND3 Cys39	Cys39 (*C-α and all side chain atoms*)
E-channel residues	ND1 (Glu143, Glu192, Glu227), ND3 (Asp66, Glu68), ND4L (Glu34, Glu70) (*C-α and all side chain atoms*)
ND6 helix π bulge	Gly61-62 (*C-α and all side chain atoms*)
Conserved charged residues in ND2, ND4 and ND5	ND2 (Glu34, Lys105, Lys263), ND4 (Glu123, Lys206, Glu378), ND5 (Glu145, Lys223, Lys392) (*C-α and all side chain atoms*)
Semi-conserved charged residues in ND2, ND4 and ND5	ND2 (His186), ND4 (His220, His293, His319), ND5 (Arg176, Asp179, His328, His332, Lys336) (*C-α and all side chain atoms*)
Charged residues associated with the transverse helix of ND5	ND2 (His112), ND4 (His213), ND5 (His230, Asp554) (*C-α and all side chain atoms*)
Loops that interrupt TMHs in ND2, ND4 and ND5	ND2 TMH4 (Gly107-Phe111) and TMH9 (Gly252-Gly259), ND4 TMH7 (Met207-Leu214) and TMH12 (Leu367-Thr372), ND5 TMH7 (Gly222-Pro231) and TMH12 (Gly382-Tyr390) (*C-α and all side chain atoms*)
ND2, ND4 and ND5 helix π bulges	ND2 (Leu131, Lys135), ND4 (Ala233, Lys237), ND5 (His248) (C-α and all side chain atoms)
Acyl groups and LYR/LYK motifs	NDUFA6 LYR motif (Leu34-Arg36), NDUFB9 LYK motif (Leu18-Lys20) (*C-α and all side chain atoms*) NDUFAB1 α/β Ser44 (*C-α and all side chain atoms*), pantetheine 4'-phosphate and acyl groups (*all atoms*)
NADPH in NDUFA9	NADPH cofactor (*all atoms*)
ADP in NDUFA10	ADP cofactor (*all atoms*)
Zinc in NDUFS6	Bound Zn^{2+}
Disulphide bonds in canonical CHCH-domain proteins	NDUFA8 (Cys35&Cys65, Cys45&Cys55, Cys87&Cys99, Cys77&Cys109), NDUFB7 (Cys58&Cys89, Cys68&Cys79), NDUFS5 (Cys32&Cys65, Cys42&Cys55) (*C-α and all side chain atoms*)
Non-canonical disulphide bonds	NDUFA11 (Cys17&Cys74, Cys94&Cys114), NDUFB10 (Cys70&Cys77, Cys106&Cys118), ND5 (Cys279) & NDUFB8 (Cys115), NDUFB11 (Cys112) & NDUFB10 (Cys148) (*C-α and all side chain atoms*)

Table 6.14 Key structural and functional features in complex I. Features were grouped as shown for distance measurements and coordinates used are indicated in italics in parenthesis.

Figure 6.11 details the positions of each feature in the structure of complex I. The redox cofactors (FMN and Fe-S clusters) are found in the hydrophilic arm along with three other bound molecules: a zinc ion in NDUFS6, ADP in NDUFA9 and NADPH in NDUFA10, which have no known function. The two copies of the acyl carrier protein (NDUFAB1) also include a bound acyl group and the LYR/K motifs of NDUFA6 and NDUFB9 interact with these subunits (Zhu *et al*., 2015). The position of the Q-channel was determined using CAVER (Chovancová *et al*., 2012) with a 1.4 Å probe and the oxygen of the hydroxyl of NDUFS2 Tyr108 used as the starting position. The 49 midline coordinates were then used in distance analyses. Figure 6.11A demonstrates the position of the Q-site at the hinge region of the matrix and membrane arm, connecting the chain of Fe-S clusters to the beginning of a network of charged residues running through the core membrane domain subunits, which is proposed to propagate conformational changes for proton translocation (Baradaran *et al*., 2013). Cysteine-39 of ND3 is an important residue, its position on a flexible loop shows altered solvent accessibility depending on the state of the enzyme. When complex I is in its deactive state this residue is exposed and can be trapped by chemical modification with fluorescent ***N***-ethyl maleimide, allowing the proportion of complex I in active and deactive states to be determined (Galkin *et al*., 2008). There are also a number of disulphides present in complex I, falling into two groups, those in canonical CHCH-domain proteins and non-canonical disulphides, including bonds between different subunits. All disulphide bonds present in the active mouse structure (Agip *et al*., 2018) were included along with three additional non-canonical disulphide bonds present in the human model: ND5 Cys279 & NDUFB8 Cys115, NDUFB11 Cys112 & NDUFB10 Cys148 and NDUFA11 Cys94 & Cys114.

Figure 6.11B highlights the key TMHs of the mitochondrially-encoded core subunits and the long transverse helix of ND5. Importantly, each of the antiporter-like subunits (ND2, 4 and 5) has a histidine on the loop of discontinuous TMH7 that points outside the subunit and may interact with the transverse helix of ND5 (ND4-His213 clamps the helix to ND4) (Agip *et al.*, 2018). These subunits also comprise two TMHs that are interrupted by loops; which have been proposed to act as channels for proton movement, as broken alpha helices often provide ion-binding sites and conformational flexibility (Efremov and Sazanov, 2011). Four TMHs also demonstrate an unusual structure, distortions – π bulges – that have been observed in resolved structures, with those in TMH8 of ND2, 4 and 5 present in all structures, although the strict definition of a π bulge is not always fulfilled (Agip *et al.*, 2018). In contrast, the ND6 π bulge on TMH3 was not present in the active mouse structure (used to generate the human model) and consequently conformational differences were observed in this region between active and deactive states – highlighting this π bulge as functionally important for catalysis (Zhu *et al*., 2016; Agip *et al.*, 2018).

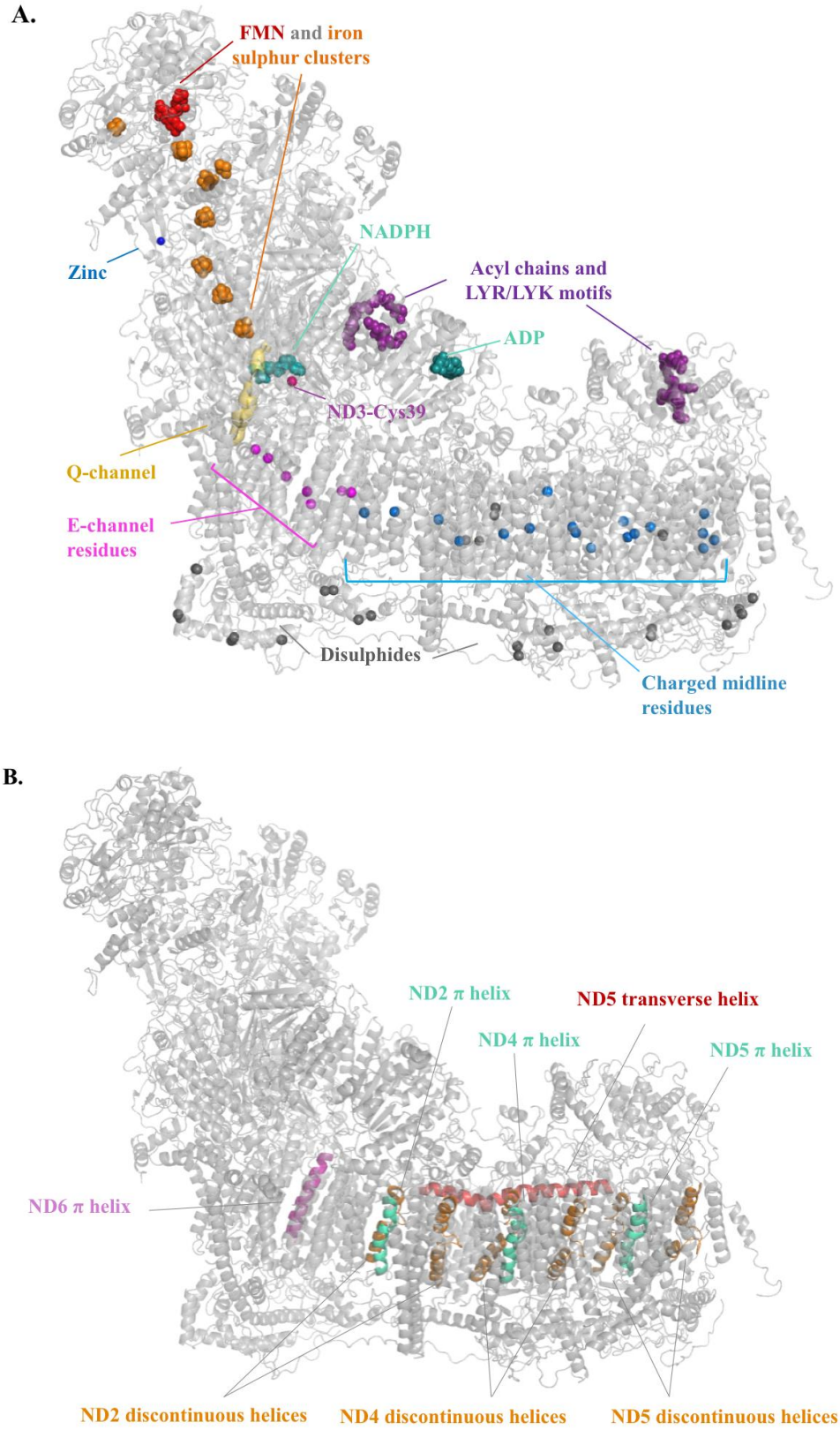


Figure 6.11 Overview of key features in complex I. A. Positions of the bound cofactors and the Q-site are indicated and the distribution of disulphide-bonded residues (black) in the supernumerary subunits. The E-channel residues of ND1, 3 and 4L are shown as pink spheres, which connect the Q-site (yellow) to the charged residues located along the midline of the three antiporter-like subunits (ND2, 4 and 5) shown in teal. Cys39 of ND3 is also highlighted; its altered solvent exposure denoted the active to deactive transition of complex I. **B.** Functionally relevant TMHs – the π bulge containing helix of ND6 (pink), the π bulge containing helix of each antiporter-like subunit (turquoise), the six TMHs interrupted by loops (orange) and the long transverse helix of ND5 (red).

These features were all chosen, although not all have assigned functional roles. Figure 6.12 provides a closer inspection of two functionally important regions – the network of charged residues and the Q-site. The E-channel charged residues were grouped togther for their proposed connecting role, linking changes within the Q-channel on quinone reduction to the proton pumping machinery of the membrane domain, Figure 6.12A, in addition to forming the fourth proton channel. Charged residues of the sequence related antiporter subunits (ND2, ND4 and ND5) are suggested to coordinate conformational changes for proton pumping and are either conserved across all three subunits or partially conserved - so were grouped accordingly (dark blue and pale blue sticks respectively in Figure 6.12A). Figure 6.12B shows the position of three key residues in the head-group binding region of the Q-channel, which were taken along with the midline coordinates of the channel as a representation of the quinone binding region. Therefore, by careful consideration of structural and biochemical information, these features were assessed with respect to the proximity of residues in the six variant classes.

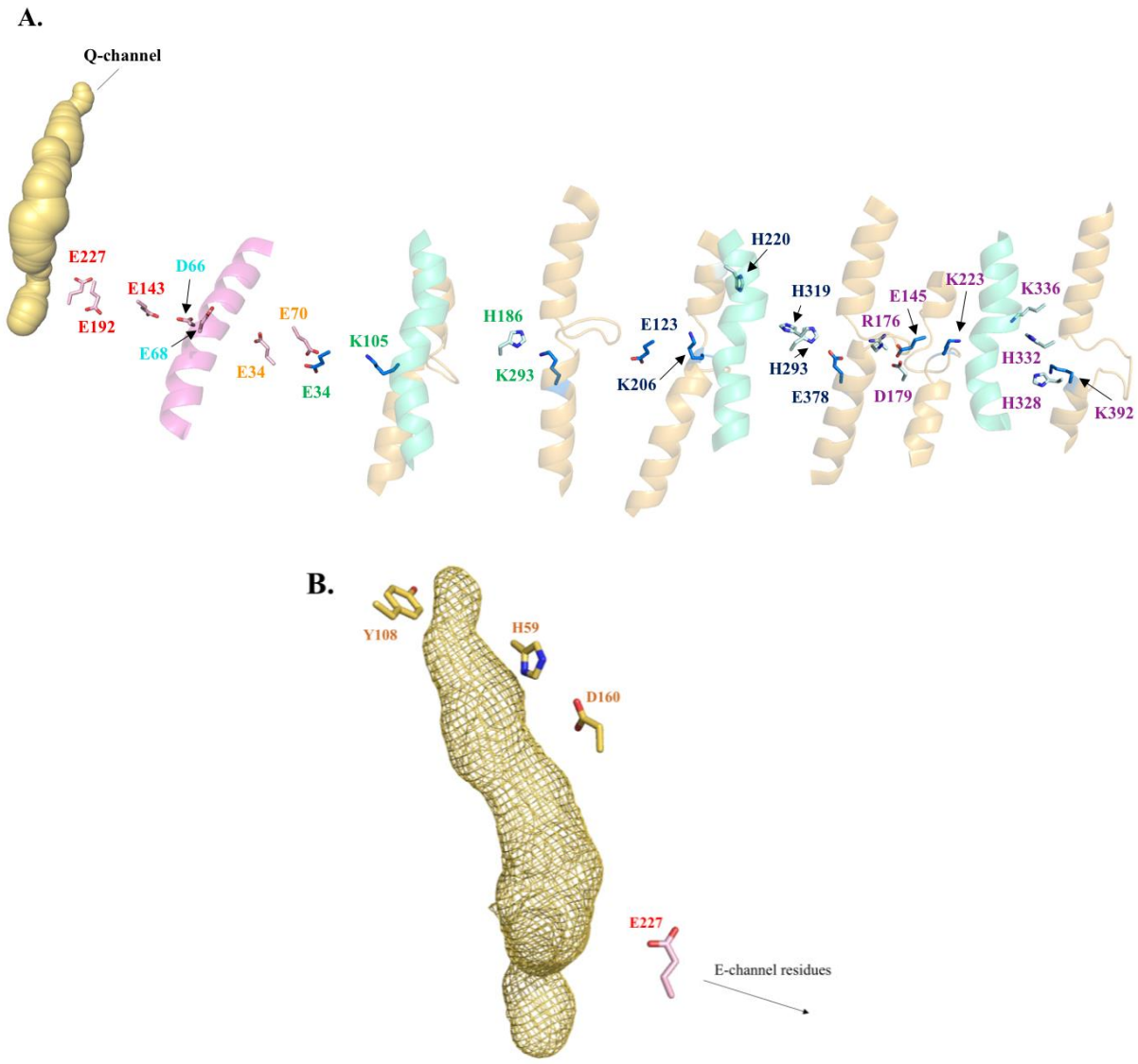


Figure 6.12 Detailed locations of residues in the charged network and the Q-site. A. E-channel residues are shown as pink sticks and residue identities coloured by subunit (ND1 in red, ND3 in cyan and ND4L in orange). The charged residues along the midline of the antiporter-like subunits are shown as dark blue sticks for conserved residues and pale blue sticks for semi-conserved residues. Residue identities are coloured by subunit (ND2 in green, ND4 in navy and ND5 in purple). Semi-transparent TMHs shown are the π bulge containing helix of ND6 (pink), π bulge containing helices of ND2, 4 and 5 (teal) and the discontinuous helices of ND2, 4 and 5 (orange). **B.** The quinone channel is shown as a yellow surface mesh and the three catalytically important residues His59, Tyr108 and Asp160 of NDUFS2 are shown as yellow sticks. The base of the Q-channel is proximal to the first of the E-channel residues, Glu227 of ND1.

6.6.5.1. The distribution of variants differs between types of key features

Figure 6.13 shows the proportions of each variant class found within 10 Å of the key features discussed, the raw data is given in Table 6.11. Figure 6.13A represents the features present in the core hydrophilic domain subunits and the adjacent mtDNA-encoded subunits (ND1, 3, 4L and 6), which do not have antiporter-like structures. Across these features there is selection for nearby variants to be pathogenic, particularly close to redox cofactors, where this is both a significant trend across the classes and significantly more class 1 mutations within 10 Å compared to class 1 polymorphisms. Additionally, compared to the distribution across the whole structure there is a higher proportion of mutations than expected. Despite this there were no variants within 10 Å of ND3-Cys39, which may suggest that this residue is intolerant to local residue changes. In contrast to the prevalence of pathogenic mutations near the features shown in Figure 6.13A, there was an opposite trend for the antiporter-like subunits. The most distinct difference was found for the loops in discontinuous TMHs, almost 20% of class 1 polymorphisms are found <10 Å from the loops of the discontinuous TMHs. This is significantly higher than for class 1 mutations, only a single variant was with 10 Å and the intermediate classes mirrored this trend, demonstrating that variants near these loops are likely to be polymorphic. It is also interesting that whilst there was no trend in variant distribution near π bulges in the anti-porter-like subunits (Figure 6.13B), there was a significant trend for proximity to the π bulge of ND6 (Figure 6.13A), with a higher proportion of variants being pathogenic. This feature is only observed in the deactive structure of complex I, thus conformational changes in this region, observed by the loss of the π bulge in the active enzyme, are fundamental for function and therefore more susceptible to changes in the local environment. The final group of features are those found in the supernumerary subunits, Figure 6.13C, and overall less than 5% of all variant classes were found near to these features (with no significant selection towards either polymorphisms or mutations).

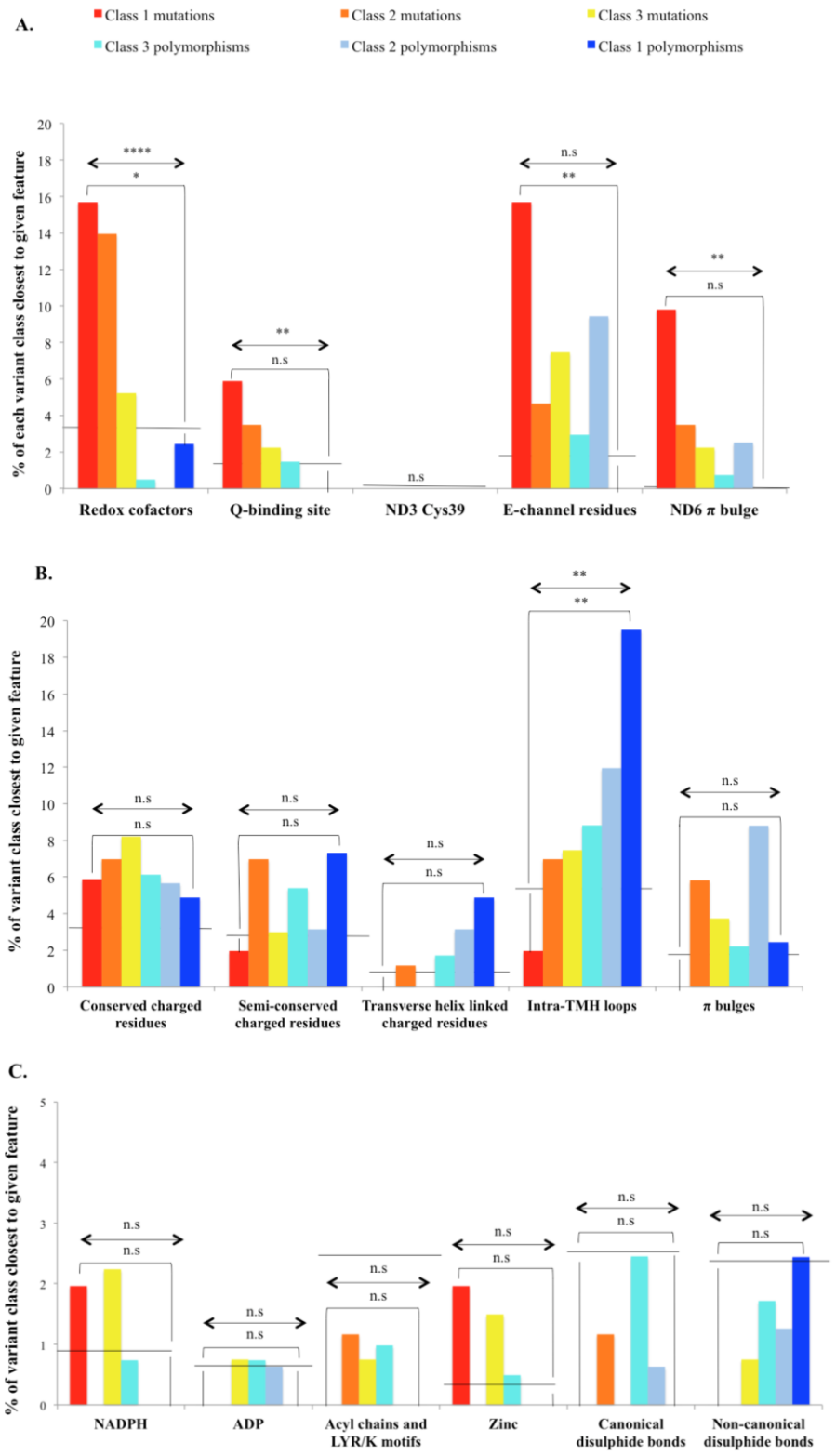


Figure 6.13 Trends in distance from key features across the classes of mutations and polymorphisms. A. Proportion of variant classes within 10 Å of features in the core nuclear-encoded subunits and neighbouring mtDNA encoded subunits ND1, 3, 4L and 6. **B.** Proportion of variant classes within 10 Å of features in the antiporter-like subunits ND2, 4 and 5. **C.** Proportion of variant classes within 10 Å of features in the supernumerary subunits. The percentage of residues within 10 Å of different features across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines.Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.15.

	Mutations			Polymorphisms			All residues
	Class 1	Class 2	Class 3	Class 3	Class 2	Class 1	
Redox cofactors	8	12	7	2	0	1	306
Q-binding site	3	3	3	6	0	0	127
ND3 Cys39	0	0	0	0	0	0	32
E-channel residues	8	4	10	12	15	0	168
ND6 π bulge	5	3	3	3	4	0	33
Conserved charged residues	3	6	11	25	9	2	282
Semi-conserved charged residues	1	6	4	22	5	3	246
Transverse helix linked charged residues	0	1	0	7	5	2	103
Intra-TMH loops	1	6	10	36	19	8	446
π bulges	0	5	5	9	14	1	164
NADPH	1	0	3	3	0	0	76
ADP	0	0	1	3	1	0	55
Acyl chains and LYR/K motifs	0	1	1	4	0	0	185
Zinc	1	0	2	2	0	0	22
Canonical disulphide bonds	0	1	0	10	1	0	211
Non-canonical disulphide bonds	0	0	1	7	2	1	183

Table 6.15 Proportion of mutations and polymorphisms found within 10 Å of key features in complex I. The number of variants from each class found in proximal to the 16 key features in complex I. The proportion of all residues in the structure present within 10 Å of each key feature is also shown, for comparision.

6.6.5.2. Grouping of complex I features by location in the structure reveals clustering of pathogenic variants

The features in Figure 6.13A, B and C were each combined to increase the sample size, demonstrating clear and opposing trends between groups A and B, representing the features present in the two core subunits groupings, Figure 6.14 and Table 6.16. This comparison reveals that key features of complex I cannot be treated as a single factor in predicting pathogenic outcome - the proximity to features in the antiporter-like subunits indicates a significant likelihood of being polymorphic, compared to features in the remaining core subunits, which demonstrate selection towards a pathogenic outcome. Distance from supernumerary subunit features did not provide predictive information. This distinctive difference suggests that residue changes near features such as the redox cofactors or the quinone-binding site are detrimental, thus manifest and are clinically identified in disease patients. However, functionally important elements of the antiporter-like subunits, particularly the discontinuous helices, are more tolerant, with variation in local residues presenting as benign mutations. This therefore implies that

TMHs are inherently less affected by changes to residue identity, as shown by the increased proportion of polymorphic variants overall. Due to the identification of the antiporter subunits as fundamental in the proton pumping function of complex I the relatively low susceptibility of these regions to detrimental mutations is unexpected. Rather it would be anticipated that mutations would cluster around all functionally important regions of the protein. This result highlights the importance of specialist knowledge of the system. Consequently these observations further demonstrate the difference between the spatially distinct oxidoreductase and proton pumping functions of the enzyme.

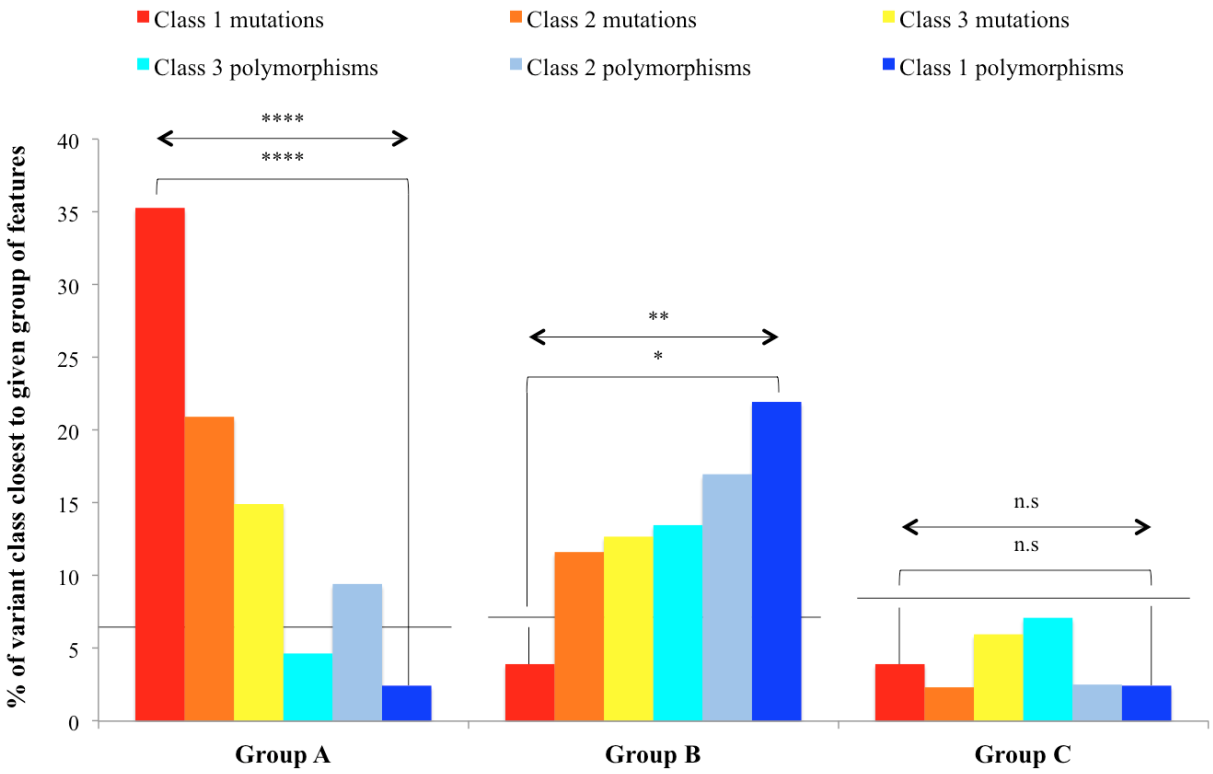


Figure 6.14 Distribution of variants found within 10 Å of different key feature groups. Key features were grouped into three sets – group A are found in the core nuclear-encoded subunits and neighbouring mtDNA-encoded subunits ND1, 3, 4L and 6 (Figure 6.12.A). Group B are features found in the antiporter-like subunits ND2, 4 and 5 (Figure 6.12.B) and group C are features of the supernumerary subunits (Figure 6.12.C). The proportions of residues within 10 Å of different features across the whole structure are indicated by the lines. The percentage expected by 'chance' was calculated from the total residues in each grouping (within the structure) compared to the total, denoted by lines. Significance of differences observed was assessed using χ^2 values from the Cochran-Armitage test on the trend (arrows) and the Fischer's Exact Test (two-tailed) between the class 1 mutations and class 1 polymorphisms - denoted * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$ and **** for $P \leq 0.0001$. Raw values are given in Table 6.16.

	Mutations			Polymorphisms			All residues
	Class 1	Class 2	Class 3	Class 3	Class 2	Class 1	
Group A	18	18	20	19	15	1	597
Group B	2	10	17	55	27	9	640
Group C	2	2	8	29	4	1	732

Table 6.16 Distribution of mutations and polymorphisms across key feature groups of complex I. The number of variants from each class found in proximal to the three groups of key features in complex I. The proportion of all residues in the structure present within 10 Å of feature group is also shown, for comparision.

Another interesting consideration is whether any of the residues identified in the key features have themselves been detected as variants in human genes. Of the 121 residues present in the 16 feature groups (Table 6.14), none were identified as polymorphic, and only five were pathogenic variants. These were the class 1 mutation ND1 Glu143 (an E-channel residue), the class 2 mutations ND3 Asp66 (an E-channel residue) and ND5 Glu145 (a conserved charged residue in the antiporter-like subunits), along with the class 3 mutations ND4 Lys 206 and ND5 Lys392 (both conserved charged residues in the antiporter-like subunits). ND1 Glu143 is a known LHON-causing variant and therefore its important function in complex I would be anticipated; mutation of residues in the E-channel network may interrupt the transfer of conformational changes induced by quinone reduction and hamper proton pumping. ND3 Asp66 as been implicated in Leigh syndrome and ND5 Glu145 in mitochondrial encephalopathy, lactic acidosis and stroke syndrome (MELAS) with differing confidence, and so again their identity as important charged residues is likely to explain the pathogenic outcome of variation. The remaining two mutations have only been reported as involved in disease: ND4 Lys206 in sporadic Creutzfelt-Jacob disease (sCJD) and ND5 Lys392 in Leigh syndrome, but their role in complex I function may strengthen their assignment. The lack of polymorphic variation in key feature residues may be informative and, combined with the low number pathogenic mutations identified, it suggests that in general these amino acids are intolerant to any variation, such that changes are embryonic lethal and incompatible with life. However, this interpretation must be made with caution, considering that not all possible residue variation has been sampled - many variants have yet to occur in nature.

6.7. Rationalising class 1 variants

Analysing different aspects of complex I individually demonstrated trends in distribution of polymorphisms and mutations, thus these criteria were then considered in combination, using class 1 mutations and polymorphisms to make an initial assessment on whether the characteristics of the residues indicated the variant outcome. Although the severity of the amino acid change was shown to be an good indicator of the variant outcome this criteria was not considered here as it required different variants at the same residue to be separated: the aim of this analysis was to gain an overview on whether sequence and structural characteristics could be used to predict the likelihood that mutation at a specific residue position would be disease-causing or benign.

Nine criteria were chosen using the previous analyses, and for each residue it was determined whether the criteria indicates pathogenicity, polymorphism or has no predictive power (colour coding in Tables 6.17 and 6.18).

- If the variant is found in a core nuclear DNA-encoded subunit this indicates pathogenicity, but if found in a mtDNA-encoded subunit this suggests a polymorphism. Supernumerary subunits were non-predictive.
- If a variant occurrs in a fully conserved residue it is more likely to be pathogenic, whereas variation in non-conserved residues has a higher propensity to be benign – the intermediate classes of conservation were non-indicative.
- Secondary structure does not aid predictions.
- If a variant is found in a surface exposed residue it is more likely to be benign, whereas changes in non-surface residues tend towards a pathogenic outcome.
- If a negatively or positively charged residue is mutated it is more likely to be pathogenic.
- Amino acids that form salt bridges are probable disease-causing variants when mutated.
- Using the distances from key features and the three groupings shown in Figure 6.13:
  - variants <10 Å from group A features (in core nuclear and neighbouring mtDNA encoded core subunits ND1, ND3, ND4L and ND6) are more likely to be pathogenic.
  - variants <10 Å from group B features (in antiporter-like subunits ND2, ND4 and ND5) are more likely to be polymorphic.
  - the distance from features in group C (supernumerary subunits) shows no pattern.

Tables 6.17 and 6.18 show the information for all class 1 mutations and polymorphisms, colour coded using the nine criteria (detailed data in Appendix 7.9). By comparing the number of pathogenic and polymorphic indicators a crude assessment was made on the outcome of variation in these residues. 39/51 class 1 mutations were demonstrated to be pathogenic, and 22 had four or more criteria demonstrating a pathogenic tendency. Another eight were assessed as likely pathogenic, with more pathogenic indicators, but only one more than the number of criteria indicating a polymorphic outcome. Two were described as ambiguous, as the outcome could not be distinguished due to equal numbers of criteria suggesting pathogenic and benign outcomes. Two also appeared likely polymorphisms, with one more criteria indicating a benign outcome compared to the pathogenic indicators: ND3 Ser34 and Ser45 are non-conserved residues in a core mtDNA-encoded subunit, with only their location at a subunit interface indicating that variation may have a pathogenic outcome. Interestingly, these two variants were also found in polymorphic classes 3 and 2, respectively, which supports the uncertainty of these as pathogenic mutations – an explanation for this may be different levels of heteroplasmy and that specific haplogroups may confer protection or susceptibility to these mutations. This was also the case for the two ambiguous residues ND1 Ala52 and Ala132, which are also found as class 3 polymorphisms. Overall this analysis demonstrated that using the combined predictive powers of different criteria 92%

of class 1 mutations were correctly identified as likely disease-causing, and none displayed characteristics that resulted in a high confidence polymorphic outcome.

The same approach was taken with class 1 polymorphisms, 18/41 were deemed polymorphic by comparing the criteria, another 14 were likely benign and seven appeared ambiguous with the same number of criteria indicating pathogenicity and neutral outcomes. The class 1 polymorphism predictions therefore agreed less with their annotations, with one likely pathogenic and one strongly pathogenic identification. NDUFV2 Val102 is found buried in a nuclear-encoded core subunit and <10 Å from an element in the group A features, all of which suggests this variant is pathogenic; no criteria highlighted it as polymorphic as it is a strongly conserved residue distant from group B key features. Therefore the annotation of this variant may be questioned; it is possible that it could cause disease in certain genetic backgrounds or with homozygous inheritance, although it was not found in any mutation class. However it is present in the wider population and causes no phenotype as class 1 polymorphisms have an allele frequency of <1%. Additionally two of the six residues assigned as ambiguous were found as class 2 (ND2 Asn150) or class 3 mutations (ND1 Tyr30), being located in mtDNA-encoded subunits where different heteroplasmy levels may affect the impact of the variant.

Overall, this analysis has shown that combining different types of criteria is a valuable approach to assessing the outcome of a variant. However this was only an initial study to compare residues with high confidence annotations: it requires further development to generate a useful predictive tool. This development will be undertaken with our collaborators, Professor Patrick Chinnery and co-workers, using computational techniques to build an equation that includes all criteria, weighted by the trends in variant distribution. Not all criteria will be treated equally (as was done above) because each criterion demonstrates different levels of correlation with variant type - thus variants patterns can be assessed accurately and complied to build an overall model biased towards criteria that have the best predictive power. The aim is to construct an equation to provide a probability score for pathogenicity, which could then be applied to every residue in complex I to map possible disease hotspot regions onto the structure. This would provide a specific tool for prediction of variants in complex I, and also provide insight into why certain variants cause disease, thus informing both clinicians and researchers, and further harnessing the information gained by structural studies on complex I.

Subunit	Identity	Subunit type	Conserv.	Secondary structure	Surface exposure	Amino acid type	Salt bridge forming	Dist. from group A features	Dist. from group B features	Dist. from group C features	Outcome
ND1	Glu24										P
ND1	Ala52										A
ND1	Ser110										LP
ND1	Gly131										LP
ND1	Ala132										A
ND1	Glu143										P
ND1	Arg195										P
ND1	Leu289										P
ND3	Ser34										LB
ND3	Ser45										LB
ND3	Ala47										LP
ND4	Arg340										P
ND4L	Val65										LP
ND5	Phe124										LP
ND5	Asp393										LP
ND6	Gly36										P
ND6	Leu60										P
ND6	Met63										P
ND6	Met64										LP
ND6	Ala72										LP
NDUFA1	Gly8										P
NDUFA1	Arg37										P
NDUFA10	Gln107										P
NDUFA9	Arg286										P
NDUFS1	Leu208										P
NDUFS1	Arg218										P
NDUFS1	Asp229										P
NDUFS1	Arg385										P
NDUFS1	Thr572										P
NDUFS1	Asp596										P
NDUFS2	Arg195										P
NDUFS2	Pro196										P
NDUFS2	Met259										P
NDUFS2	Ser380										P
NDUFS2	Ala393										P
NDUFS3	Thr109										P
NDUFS3	Arg163										P
NDUFS6	Cys87										P
NDUFS7	Val84										P
NDUFS8	Glu29										P
NDUFS8	Arg43										P
NDUFS8	Pro45										P
NDUFS8	Pro51										P
NDUFS8	Arg68										P
NDUFS8	Arg104										P
NDUFS8	Ala125										P
NDUFV1	Ala97										P
NDUFV1	Pro102										P
NDUFV1	Cys186										P
NDUFV1	Glu194										P
NDUFV1	Thr403										P

Table 6.17 Residue criteria of class 1 mutations colour coded by pathogenic or polymorphic indicators. The 51 highest confidence mutations are given along with nine criteria analysed to predict variant outcome. Red boxes denote pathogenic likelihood and blue suggest the variant is benign, uncoloured boxes are non-predictive criteria (including secondary structure as a whole). The outcome was determined by the difference in the number of pathogenic and benign indicators – a difference of: >+1 were assigned pathogenic (P), +1 as likely pathogenic (LP), no difference as ambiguous (A), -1 are likely benign (LB), <-1 are benign (B). Column heading abbreviations: conservation (Conserv.), distance from group A features (Dist. from group A features), distance from group B features (Dist. from group B features) and distance from group C features (Dist. from group C features).

Subunit	Identity	Subunit type	Conserv.	Secondary structure	Surface exposure	Amino acid type	Salt bridge forming	Dist. from group A features	Dist. from group B features	Dist. from group C features	Outcome
ND1	Ala4										B
ND1	Ile81										A
ND1	Tyr30										A
ND1	Thr67										A
ND1	Thr164										LB
ND1	Thr240										LB
ND1	Tyr304										LB
ND1	Ile309										LB
ND2	Val8										LB
ND2	Thr119										B
ND2	Asn150										A
ND2	Val193										LB
ND2	Leu237										LB
ND2	Ala331										B
ND2	Phe325										B
ND3	Val88										LB
ND3	Thr114										B
ND4	Pro140										B
ND4	Ala404										LB
ND4L	Met47										LB
ND5	Thr8										LB
ND5	Val24										B
ND5	Ile257										LB
ND5	Met314										LB
ND5	Ala458										B
ND5	Ala475										B
ND5	Ile482										LB
ND5	Tyr485										B
ND5	Ser531										B
ND5	Thr533										B
ND5	Thr544										B
ND5	Ser515										B
ND5	Leu555										B
ND6	Asn119										B
ND6	Ile166										A
NDUFA7	Lys47										A
NDUFA9	Pro259										A
NDUFB9	Pro145										B
NDUFC2	Leu46										B
NDUFS2	Pro319										LP
NDUFV2	Val102										P

Table 6.18 Residue criteria of class 1 polymorphisms colour coded by pathogenic or polymorphic indicators. The 41 highest confidence polymorphisms are given along with nine criteria analysed to predict variant outcome. Red boxes denote pathogenic likelihood and blue suggest the variant is benign, uncoloured boxes are non-predictive criteria (including secondary structure as a whole). The outcome was determined by the difference in the number of benign and pathogenic indicators – a difference of: >+1 were assigned benign (B), +1 as likely benign (LB), no difference as ambiguous (A), -1 are likely pathogenic (LP), <-1 are pathogenic (P). Column heading abbreviations: conservation (Conserv.), distance from group A features (Dist. from group A features), distance from group B features (Dist. from group B features) and distance from group C features (Dist. from group C features).

6.8. Conclusions

The work in this chapter aimed to increase our ability to predict the outcome of variants in human complex I genes by combining recent structural information with more conventional analyses of primary sequence and secondary structure. Generating a human model from the best resolution mammalian structures available allowed spatial mapping of all variants collated from online databases. The model was then used to create a comprehensive table of all residues in the structure along with details of eight characteristics: subunit type, encoding genome, conservation, secondary structure, residue environment, amino acid type, participation in salt bridges and proximity to key features. From analysis of this data regarding the distribution of pathogenic and polymorphic variants a number of insights were made, exemplifying the importance of using all available information when interpreting the outcome of a variant and supporting the call for bespoke tools that are gene/protein specific.

Firstly, conservation in complex I subunits was shown to be a strong predictive criterion, as mutations in conserved residues were more likely to cause disease than those in residues that have diverged in identity during mammalian evolution. Conservation was also closely linked to the severity of the amino acid change: drastic changes in residue identity were pathogenic but mild changes could be tolerated and were more likely to present as polymorphisms. In contrast, secondary structure alone did not help to identify pathogenic variants – trends seen in TMHs were linked to the mtDNA-encoded subunits, which were generally less well conserved and comprised a high proportion of TMHs, being found in the hydrophobic arm of the enzyme. Therefore although polymorphisms appeared to segregate into TMHs, secondary structure was not an independent predictor. This outcome may represent a distinction between complex I and other proteins, where secondary structure elements display greater independence; hence specific considerations must be made depending on the protein.

Complex I is unusual as it comprises subunits encoded by two genomes. Variation is greater in the mtDNA-encoded subunits (26% of residues compared to 5% of nuclear-encoded subunits) and reflects the difference in mutation rates of the two genomes, something that will not affect other proteins and complexes with a single encoding genome. On top of this complex I subunits can be separated into core and supernumerary subunits, those that are present in the bacterial enzyme and have a defined function, compared to more recent evolutionary additions that appear to act as scaffolding surrounding the core subunits. Variants were found to distribute differently between these two groups and therefore this represents another characteristic of complex I to be considered that may not be applicable to other proteins and could be overlooked by broad prediction tools.

A primary aim of this chapter was to utilise information that can only be obtained from protein structure; this information was scrutinised and clearly showed that criteria such as residue environment, participation in stabilising bonds (such as salt bridges) and proximity to functional regions increase the reliability of interpreting the outcome of a variant. Detrimental complex I variants clustered away from

the surface, perhaps as a result of its large multi-subunit structure, indicating that the most important interactions occur inside the complex. However, other proteins may show the inverse trend, such as those with chaperoning function, where surface association with partner proteins is critical. Additionally, as discussed above, full consideration of functionally and structural relevant features is required, with high-resolution structures enabling distances of all residues from these regions to be measured and analysed for trends. It was striking that not all of the features that are widely agreed to participate in the mechanism of complex I oxidoreduction and proton pumping are equally susceptible to mutations in nearby residues. Cofactor binding sites and the long quinone binding cavity are likely to be detrimentally affected by variation in the local environment - a high proportion of mutations was found <10 Å away but levels of polymorphisms were low. Conversely, the broken TMHs of the antiporter subunits demonstrated greater resilience; variation near these features tended to result in polymorphisms. Complex I incorporates many important sites, due to its large size, number of subunits and energy transferring function, and this work shows that these should be addressed separately to ensure specific properties, that may help identify pathogenic hotspots, are not overlooked.

Overall, this work provides a step towards development of pathogenicity prediction tools that include a wide range of protein related criteria, and exemplifies how structural research can be harnessed to aid clinicians when assessing human gene variants. However, like all prediction analyses that use existing pathogenic and polymorphic annotations, the quality of the analysis is reliant on the databases used and misinterpreted entries affect the robustness of predictions. Nevertheless variant annotation is often updated and corrected by greater sequencing of patients and analysis of controls. New databases are being released that significantly increase the data available, such as the recent genome aggregation database (gnomAD) – a development of ExAc - comprising 123,136 exome sequences and 15,496 whole genome sequences for polymorphic variant analysis. Also, whilst comparison of pathogenic and polymorphic variants provides a strong basis for determining residue susceptibly, these variants do not cover the entire spectrum of changes. The most severe mutations will never be accounted for because these changes are embryonic lethal, thus residues that display very low variation, particularly with complete lack of polymorphisms, may in fact be vital and variation at these positions is never detected. Furthermore, not every possible variation will have occurred, further decreasing the number of mutations observed. The main complication in human disease is the dual inheritance of alleles; disease-causing mutations may be recessive and require both mutated gene copies, or present as compound heterozygotes, whereby two different mutant alleles are present. Therfore assigning the causal variant is difficult, and both mutations may contribute to the phenotype. In addition mitochondrial heteroplasmy may vary, affecting the outcome. Therefore for the most robust analysis of disease variants biochemical analysis is also required. Models such as *Y. lipolytica* (described in Chapter 5) present malleable systems to study individual mutations, thereby predictions and clinical observations may be used to inform experiments that in turn provide greater information about the effect of the residue change on complex I activity and overall mitochondrial function.

Chapter 7

Appendices

Appendix 7.1. SILAC data for 143B cells grown in riboflavin depleted media (5 and 13 days) and 143B CyWT cells grown in riboflavin rich media

Complex I subunits											
Gene names	**Normalised Log_2 Forward Ratio H/L (8 days)**	**Normalised Log_2 Forward Ratio H/L (13 days)**	**Unique peptides (8 days)**	**Unique peptides (13 days)**	**Count (8 days)**	**Count (13 days)**	**Ratio H/L variability [%] (8 days)**	**Ratio H/L variability [%] (13 days)**	**Sequence coverage [%] (5 days)**	**Sequence coverage [%] (13 days)**	**Molecular weight [kDa]**
NDUFS1	2.06	1.80	11	22	11	29	18.95	27.19	20.4	41.7	79.47
NDUFV2	1.97	1.55	2	5	3	9	9.50	69.24	14.3	24.2	25.43
NDUFV1	1.94	1.39	10	17	9	18	30.94	49.00	23.0	39.4	50.05
NDUFS6	2.06	1.07	2	5	3	7	38.16	69.75	21.8	41.9	13.71
NDUFA7	1.74	0.82	2	6	2	6	62.00	94.92	16.8	47.8	12.55
NDUFA12	-	0.80	-	7	-	9	-	102.64	-	52.4	17.11
NDUFA6	-	0.69	-	3	-	5	-	19.66	-	27.3	15.14
NDUFB9	-0.14	0.45	6	8	9	15	9.12	42.36	48.8	42.9	20.38
NDUFA8	-	0.44	-	3	-	3	-	47.72	-	20.9	20.11
NDUFB8	-0.54	0.40	2	3	2	5	5.93	12.76	10.8	17.2	21.77
NDUFA3	-	0.28	-	1	-	3	-	3.94	-	13.1	9.28
NDUFB4	-0.23	0.26	4	5	5	9	11.20	34.97	41.9	41.1	15.21
NDUFA10	-0.12	0.24	9	13	11	22	50.02	84.41	25.7	32.2	40.68
NDUFA2	-	0.22	-	3	-	3	-	45.12	-	21.2	10.92
NDUFA13	-0.04	0.21	2	8	2	10	16.54	64.34	14.6	43.1	16.70
NDUFB7	-	0.18	-	4	-	8	-	21.90	-	26.3	16.40
NDUFB11	0.10	0.14	2	2	4	7	3.20	16.76	22.2	22.2	17.32
NDUFB5	0.18	0.13	3	4	3	11	5.45	29.68	13.0	23.3	20.04
NDUFS7	-0.07	0.11	2	5	2	7	2.18	13.69	12.6	18.7	19.77
NDUFS3	0.03	0.05	3	8	5	11	39.28	6.16	19.5	36.3	22.22
NDUFA9	-0.69	-0.02	4	14	4	17	44.32	64.96	11.8	46.4	38.44
MT-ND3	-	-0.03	0	1	-	2	-	1.39	-	28.3	6.09
NDUFS2	-0.04	-0.05	14	22	14	24	20.91	59.77	39.4	54.7	46.40
NDUFB10	0.10	-0.06	4	7	6	15	7.28	51.34	21.7	51.6	19.26
NDUFS8	-	-0.10	0	1	-	4	-	42.69	-	23.3	19.21
MT-ND5	-	-0.12	1	3	-	4	-	27.01	-	8.6	53.19
NDUFA5	-	-0.12	1	3	-	3	-	4.89	-	50.0	7.84
NDUFB3	-	-0.13	1	3	-	8	-	99.08	-	29.2	7.59

NDUFB6	0.16	-0.24	2	5	3	8	21.03	90.98	15.6	32.8	15.49
NDUFS5	-	-0.32	1	3	-	5	-	29.53	-	32.1	12.52
NDUFC2	-0.40	-0.33	1	5	2	6	60.40	150.65	10.1	31.9	14.19
NDUFB1	0.04	-1.25	2	3	3	7	12.68	50.07	32.8	34.5	6.96

Mitoflavoproteins											
Gene names	**Normalised Log_2 Forward Ratio H/L (8 days)**	**Normalised Log_2 Forward Ratio H/L (13 days)**	**Unique peptides (8 days)**	**Unique peptides (13 days)**	**Count (8 days)**	**Count (13 days)**	**Ratio H/L variability [%] (8 days)**	**Ratio H/L variability [%] (13 days)**	**Sequence coverage [%] (5 days)**	**Sequence coverage [%] (13 days)**	**Molecular weight [kDa]**
SQRDL	2.86	4.94	13	18	11	8	21.06	23.35	34.9	53.7	44.15
DLD	3.14	3.06	7	7	5	9	18.46	40.13	19.2	23.1	48.90
SDHA	2.14	2.38	11	21	10	24	19.39	28.80	23.6	45.8	72.69
AIFM1	1.56	2.11	14	21	14	28	22.36	29.93	26.4	41.6	66.90
ACADM	1.79	1.56	3	6	3	6	43.53	61.54	10.1	22.9	42.43
GPD2	1.11	1.52	23	35	20	43	25.80	71.56	36.9	54.5	77.25
NDUFV1	1.94	1.39	10	17	9	18	30.94	49.00	23.0	39.4	50.05
ACADVL	1.38	1.25	17	21	13	19	32.58	54.23	35.7	34.5	70.39
ETFA	0.32	0.84	3	5	5	11	5.58	15.00	25.0	49.6	24.90
FOXRED1	-	0.84	0	6	-	4	-	7.22	-	16.5	53.81
L2HGDH	0.22	0.64	3	3	3	2	31.82	9.62	7.0	8.6	48.48
IVD	-	0.45	1	1	-	4	-	22.03	-	13.0	46.32
GSR	-	0.33	0	4	-	4	-	62.05	-	11.7	56.26
DHODH	0.72	0.15	4	18	4	18	32.10	45.80	15.2	47.3	42.87
ACAD9	0.26	-0.07	16	30	15	34	11.25	42.44	31.6	45.2	68.76
CYB5R3	-0.50	-0.59	6	7	5	5	6.32	22.02	24.4	30.6	33.24
COQ6	-0.90	-0.89	3	7	3	4	43.69	6.79	9.0	20.3	50.80

Appendix 7.2. Complexome data for mitochondrial membrane and matrix fractions for 143B CyWT cells and conditioned RFK KD 143B cells

7.2A Table of proteins used in mass estimation of 143B CyWT membrane sample (Calibration curve Figure 3.20A)

143B CyWT membrane fraction calibration proteins				
Protein	**Mass (kDa)**	**Native state**	**Total Mass (kDa)**	**Gel slice**
Mitochondrial ribosome (39S + 28S)	-	39S + 28S subunits	2900	10
Complex I (NADH:ubiquinone oxidoreductase)	1000	monomer	1000	19.5
Complex V (ATP synthase)	770	monomer	770	24
Complex III (Cytochrome bc-1 complex)	245	dimer	490	28
NADP transhydrogenase	114	dimer	228	35
Complex IV (cytochrome c oxidase)	200	monomer	200	38
SLC25A13 (citrin)	74	monomer	74	39
SLC25A12 (aralar 1)	75	monomer	75	39
Complex II (succinate dehydrogenase)	130	monomer	130	44
Dihydrolipoamide dehydrogenase	54	monomer	54	45
Citrate synthase	52	dimer	104	50
Electron transfer ubiquinone oxidoreductase	69	monomer	69	52
Choline dehydrogenase	65	monomer	65	52
SLC25A23 (calcium binding carrier protein)	52	monomer	52	53
Voltage dependent ion gated channel 1	31	monomer	31	53
Cytochrome c	12	monomer	12	55
ADP/ATP translocase (adenine nucleotide translocase)	30	monomer	30	55
2-oxoglutarate/malate carrier protein	34	dimer	68	57
Sulphide:quinone oxidoreductase	49	monomer	49	57
SLC25A3 (phosphate carrier protein)	40	monomer	40	57
Ubiquinone biosynthesis protein COQ9	37	monomer	37	62
Adenylate kinase 2	26	monomer	26	63

7.2B Table of proteins used in mass estimation of RFK KD membrane sample (Calibration curve Figure 3.20B)

RFK KD membrane fraction calibration proteins				
Protein	**Mass (kDa)**	**Native state**	**Total Mass (kDa)**	**Gel slice**
Mitochondrial ribosome (39S + 28S)	-	39S + 28S subunits	2900	10
Complex I (NADH:ubiquinone oxidoreductase)	1000	monomer	1000	21
Propionyl-CoA carboxylase (α and β subunits)	α -80 and β - 58	heterododecamer	820	21
Complex V (ATP synthase)	770	monomer	770	25
Complex III (Cytochrome bc-1 complex)	245	dimer	490	30
NADP transhydrogenase	114	dimer	228	35

Glutamate dehydrogenase	61	hexamer	366	36
Polyribonucleotide nucleotidyltransferase 1	86	trimer	258	38
Complex IV (cytochrome c oxidase)	200	monomer	200	40
SLC25A13 (citrin)	74	monomer	74	41
SLC25A12 (aralar 1)	75	monomer	75	41
Methylmalonyl-CoA mutase	83	dimer	166	42
Complex II (succinate dehydrogenase)	130	monomer	130	47
Aconitate hydratase	85	monomer	85	48
Dihydrolipoamide dehydrogenase	54	monomer	54	50
D-beta-hydroxybutyrate dehydrogenase	38	tetramer	152	50
Citrate synthase	52	dimer	104	55
SLC25A23 (calcium binding carrier protein)	52	monomer	52	58
Thymidylate synthase	36	dimer	72	59
Electron transfer ubiquinone oxidoreductase	69	monomer	69	59
ADP/ATP translocase (adenine nucleotide translocase)	30	monomer	30	61
Dihydroorotate dehydrogenase	43	monomer	43	62
Cytochrome c	12	monomer	12	63
2-oxoglutarate/malate carrier protein	34	dimer	68	63
L-2-hydroxyglutarate dehydrogenase	50	monomer	50	63
SLC25A32 (folate transporter/carrier)	35	monomer	35	64

7.2C Complexome data for mitochondrial membrane from 143B CyWT cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-ND1	8.49	2	35.6	29.44
MT-ND2	7.49	2	38.9	34.32
MT-ND4	14.16	6	51.5	93.57
MT-ND5	11.61	3	67.0	153.12
MT-ND6	8.05	1	18.6	60.08
NDUFA10	25.07	7	40.7	126.43
NDUFA11	31.91	3	14.8	79.96
NDUFA12	35.17	4	17.1	83.52
NDUFA13	43.06	8	16.7	362.04
NDUFA2	48.48	5	10.9	259.59
NDUFA3	36.90	2	9.3	49.40
NDUFA5	30.17	4	13.5	142.12
NDUFA6	33.77	5	17.9	155.25
NDUFA7	62.83	7	12.5	108.97
NDUFA8	21.51	3	20.1	84.66
NDUFA9	47.75	15	42.5	253.84
NDUFB1	34.48	3	7.0	107.39

NDUFB10	43.60	7	20.8	251.21
NDUFB11	28.76	2	17.3	70.61
NDUFB3	28.57	3	11.4	73.44
NDUFB4	36.43	4	15.2	115.55
NDUFB5	18.52	4	21.7	93.34
NDUFB6	42.19	4	15.5	110.99
NDUFB7	38.69	3	16.4	243.46
NDUFB8	20.97	3	21.8	119.90
NDUFB9	21.23	3	21.8	98.48
NDUFC2	31.09	4	14.2	58.46
NDUFS1	45.12	24	79.4	877.24
NDUFS2	33.69	12	52.5	326.21
NDUFS3	46.59	10	30.2	282.20
NDUFS4	34.86	4	20.1	105.79
NDUFS5	34.91	4	12.5	173.83
NDUFS6	37.10	5	13.7	41.67
NDUFS7	18.31	5	23.5	318.60
NDUFS8	23.81	5	23.7	95.31
NDUFV1	18.75	9	50.8	182.99
NDUFV2	37.35	8	27.4	438.09
NDUFV3	14.81	1	11.9	16.69

Complex II subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
SDHA	31.63	16	72.6	564.76
SDHB	33.93	11	31.6	414.76
SDHC	4.73	1	18.6	45.56
SDHD	10.69	1	17.0	35.8

Complex III subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-CYB	2.37	1	42.7	97.43
UQCR10	38.10	2	7.3	171.50
UQCR11	21.43	1	6.6	15.56
UQCRB	47.75	6	13.5	249.50
UQCRC1	32.29	12	52.6	1091.78
UQCRC2	40.18	13	48.4	1934.88
UQCRFS1	44.89	11	29.6	674.08
UQCRH	28.57	1	10.7	50.74
UQCRQ	56.10	5	9.9	64.32

Complex IV subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
COX4I1	44.97	7	19.6	721.04
COX5A	32.67	5	16.8	717.80
COX5B	41.09	6	13.7	388.99
COX6A1	16.51	1	12.1	42.90
COX6B1	19.77	3	10.2	116.80
COX6C	66.67	7	8.8	379.16
COX7A2	56.63	2	9.4	722.64
COX7A2L	70.18	4	12.6	56.33
COX7C	25.40	1	7.2	39.91
COX8A	13.04	1	7.6	66.75
MT-CO1	7.60	3	57.0	76.99
MT-CO2	27.75	5	25.5	744.26

Complex V subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
ATP5A1	51.72	25	59.7	13821.39
ATP5B	70.70	25	56.5	15750.88
ATP5C1	37.92	10	33.0	1034.04
ATP5D	13.69	2	17.5	290.75
ATP5E	31.37	2	5.8	93.60
ATP5F1	56.25	16	28.9	1313.04
ATP5H	75.78	12	18.5	721.14
ATP5I	62.32	5	7.9	332.53
ATP5J	46.30	4	12.6	334.23
ATP5J2	25.53	2	10.9	515.13
ATP5L	89.32	6	11.4	683.56
ATP5O	55.40	10	23.3	1701.88
MT-ATP6	4.42	1	24.8	222.46
MT-ATP8	36.76	2	8.0	29.33
ATPIF1	7.55	1	12.2	19.88

Complex I assembly factors				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
ACAD9	30.6	17	68.7	357.26
COA1	13.01	2	16.7	23.77
ECSIT	28.31	8	49.1	85.24
NDUFAF1	32.11	7	37.7	169.02
NDUFAF2	65.68	9	19.8	99.48
NDUFAF3	41.85	6	20.3	93.32
NDUFAF4	29.71	4	20.3	51.81
TIMMDC1	29.47	7	32.2	63.85

7.2D Complexome data for mitochondrial membrane from conditioned RFK KD 143B cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-ND1	8.81	3	35.6	234.16
MT-ND2	2.59	1	38.9	42.19
MT-ND4	13.94	5	51.5	132.58
MT-ND5	9.29	5	67.0	382.40
MT-ND6	8.05	1	18.6	111.95
NDUFA10	25.07	10	40.7	468.25
NDUFA11	21.28	2	14.8	167.94
NDUFA12	80.69	11	17.1	686.00
NDUFA13	63.89	11	16.7	1091.22
NDUFA2	48.48	5	10.9	389.63
NDUFA3	50.00	3	9.3	154.96
NDUFA5	36.21	5	13.5	353.47
NDUFA6	33.77	5	17.9	347.09
NDUFA7	61.06	9	12.5	619.23
NDUFA8	21.51	4	20.1	282.08
NDUFA9	57.82	21	42.5	841.02
NDUFAB1	11.54	1	17.4	103.64
NDUFB1	36.21	3	7.0	263.46
NDUFB10	53.49	12	20.8	1157.36
NDUFB11	35.29	2	17.3	96.95
NDUFB2	8.57	1	12.1	30.73
NDUFB3	37.76	5	11.4	98.35
NDUFB4	49.61	6	15.2	213.48
NDUFB5	28.04	6	21.7	363.66

NDUFB6	58.59	5	15.5	355.40
NDUFB7	40.15	5	16.4	1242.45
NDUFB8	32.26	6	21.8	83.52
NDUFB9	49.72	8	21.8	356.49
NDUFC2	50.42	8	14.2	60.42
NDUFS1	50.21	32	79.4	33149.44
NDUFS2	45.79	16	52.5	716.48
NDUFS3	51.52	14	30.2	1142.28
NDUFS4	29.14	6	20.1	391.13
NDUFS5	34.91	4	12.5	138.65
NDUFS6	38.71	4	13.7	440.87
NDUFS7	23.47	6	23.5	639.56
NDUFS8	27.62	5	23.7	499.44
NDUFV1	35.78	16	50.8	496.12
NDUFV2	40.16	10	27.4	1138.18
NDUFV3	14.81	1	11.9	58.81

Complex II subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
SDHD	11.32	2	17.0	66.12
SDHA	45.48	22	72.6	904.21
SDHB	40.36	13	31.6	527.71
SDHC	14.20	1	18.6	67.75

Complex III subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-CYB	5.00	2	42.7	79.79
UQCRC1	42.71	20	52.6	2412.52
UQCR11	21.43	1	6.6	82.33
UQCRC2	45.25	17	48.4	1948.11
UQCRH	28.57	1	10.7	163.22
UQCRB	56.76	9	13.5	525.63
UQCRQ	95.12	9	9.9	343.21
UQCR10	39.68	3	7.3	164.86
UQCRFS1	60.95	14	29.6	836.57

Complex IV subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-CO1	7.60	3	57.0	448.88
MT-CO2	27.75	5	25.5	1249.18

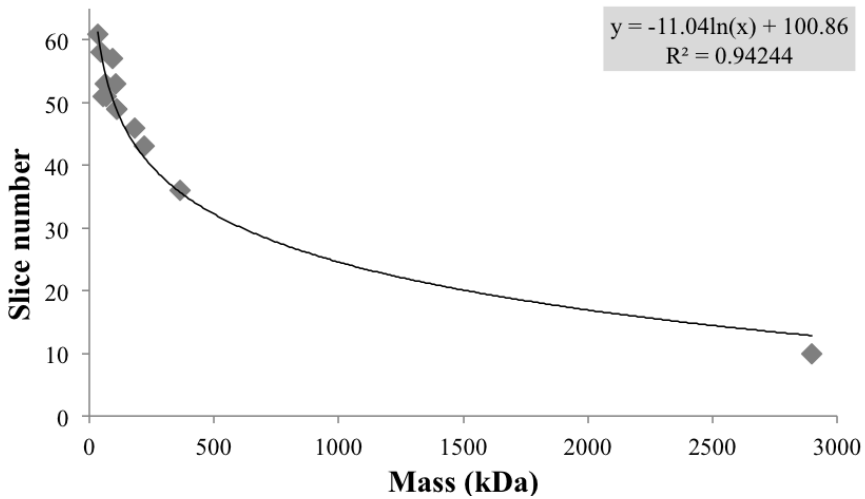
MT-CO3	8.81	2	29.9	40.96
COX4I1	55.03	12	19.6	1114.07
COX5A	38.67	7	16.8	780.93
COX5B	51.16	8	13.7	451.38
COX6A1	35.78	2	12.1	50.39
COX6B1	19.77	3	10.2	97.63
COX6C	76.00	10	8.8	659.18
COX7A2	56.63	3	9.4	241.17
COX7C	28.57	2	7.2	44.58
COX8A	13.04	1	7.6	125.73

Complex I assembly factors				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
ACAD9	42.67	27	68.7	899.87
ECSIT	41.53	11	49.1	590.82
FOXRED1	2.67	1	53.8	81.00
NDUFAF1	42.20	11	37.7	383.67
NDUFAF2	39.05	5	19.8	59.88
NDUFAF3	25.00	4	20.3	257.55
NDUFAF4	43.43	8	20.3	295.57
TIMMDC1	44.21	10	32.2	332.44
TMEM126B	29.57	4	25.9	57.22
TMEM70	11.54	2	29.0	77.05

7.2E Complexome data for mitochondrial matrix from 143B CyWT cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
NDUFA2	10.10	1	10.9	10.00
NDUFS1	9.77	3	79.4	42.56
NDUFV1	18.53	6	50.8	18.58
NDUFV2	24.10	4	27.4	21.50

7.2F Calibration curve for mass estimation of RFK KD matrix sample.



7.2G Table of proteins used in mass estimation of RFK KD matrix sample.

RFK KD matrix fraction calibration proteins				
Protein	**Mass (kDa)**	**Native state**	**Total Mass**	**Gel slice**
Mitochondrial ribosome (39S + 28S)	-	39S + 28S subunits	2900	10
Glutamate dehydrogenase	61	hexamer	366	36
Fumarate hydratase	55	tetramer	220	43
Acetyl-CoA acetyltransferase	180		180	46
Dynamin-like 120 kDa protein	112	monomer	112	49
Citrate lyase subunit beta-like protein	37	trimer	111	49
Acyl-CoA dehydrogenase family member 9	51	monomer	69	51
Dihydrolipoamide dehydrogenase	54	monomer	54	51
Electron transfer protein (α and β subunits)	α - 35 and β - 28	heterodimer	63	53
Citrate synthase	52	dimer	104	53
Kynurenine/alpha-aminoadipate aminotransferase	47	dimer	94	57
Medium chain specific acyl-CoA dehydrogenase	46	monomer	46	58
39S ribosomal protein L1	36	monomer	36	61

7.2H Complexome data for mitochondrial matrix from conditioned RFK KD 143B cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
NDUFA13	9.03	1	16.7	19.21
NDUFA2	31.31	2	10.9	112.03
NDUFA5	17.24	2	13.5	25.88
NDUFAB1	11.54	2	17.4	74.46
NDUFS1	36.18	19	79.4	49.14
NDUFS2	2.59	1	52.5	27.92
NDUFS3	16.29	3	30.2	30.06
NDUFS8	9.05	1	23.7	19.47
NDUFV1	23.28	8	50.8	315.38
NDUFV2	38.96	9	27.4	585.16

Complex II subunits and assembly factors				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
SDHA	43.83	19	72.6	1333.16
SDHAF2	16.27	2	19.6	123.43
SDHB	39.64	11	31.6	249.77

Appendix 7.3 SILAC data for conditioned RFK KD 143B cells and riboflavin rich 143B CyWT cells

Complex I subunits									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
MT-ND1	1.38	0.75	3	11	138.00	11.48	3	8.8	35.66
MT-ND3	0.77	0.52	5	4	25.74	12.62	1	13.0	13.19
MT-ND4	-0.28	-0.02	3	4	28.81	17.86	4	10.9	51.58
MT-ND5	0.45	0.60	4	6	21.22	6.27	4	8.0	67.03
NDUFA10	0.49	0.84	40	45	22.44	23.12	21	47.6	40.75
NDUFA11	1.00	1.13	4	10	17.00	8.52	5	41.8	14.85
NDUFA12	2.68	1.59	6	24	15.78	19.89	10	82.8	17.11
NDUFA13	0.43	0.79	17	24	34.88	8.13	7	52.8	16.70
NDUFA2	1.60	1.74	6	11	61.20	38.33	5	42.4	10.92
NDUFA3	0.48	0.68	7	12	14.65	4.80	3	50.0	9.28
NDUFA5	0.02	-0.02	14	18	22.89	14.26	7	49.1	13.46

NDUFA6	1.65	1.31	6	9	13.68	16.22	4	33.6	15.14
NDUFA7	2.93	1.79	6	24	31.37	18.81	9	69.0	12.55
NDUFA8	1.35	0.86	10	20	36.24	13.33	8	47.7	20.11
NDUFA9	0.79	0.98	28	42	20.46	21.40	21	60.2	42.51
NDUFAB1	0.52	0.83	3	5	39.55	8.24	4	21.2	17.42
NDUFB1	-0.06	0.37	9	10	30.13	29.01	3	34.5	6.96
NDUFB10	0.43	0.63	56	62	14.68	14.25	17	61.6	20.78
NDUFB11	0.23	0.44	18	20	12.22	13.07	6	45.8	17.32
NDUFB2	0.84	0.82	2	5	18.95	15.18	2	11.5	9.39
NDUFB3	0.64	0.74	16	19	28.39	14.79	5	32.7	11.40
NDUFB4	0.38	0.76	11	21	53.28	22.23	8	50.4	15.21
NDUFB5	0.38	0.63	11	17	6.50	11.37	6	35.0	15.67
NDUFB6	0.22	0.65	12	27	20.35	13.98	7	51.6	15.49
NDUFB7	0.98	0.72	11	17	21.26	8.11	6	48.2	16.40
NDUFB8	1.01	0.74	18	20	32.31	9.87	6	38.7	18.48
NDUFB9	0.23	0.62	46	29	21.08	7.42	12	69.3	21.83
NDUFC2	0.72	0.53	15	25	29.80	14.52	9	45.4	14.19
NDUFS1	2.18	1.54	24	72	40.72	23.60	37	63.8	79.47
NDUFS2	0.11	0.15	43	62	19.23	16.38	24	57.0	52.55
NDUFS3	0.05	-0.06	32	63	22.89	14.25	18	58.0	30.24
NDUFS4	2.30	1.88	2	10	86.93	23.27	5	25.7	20.11
NDUFS5	1.11	0.68	5	10	25.91	10.41	6	39.6	12.52
NDUFS6	2.62	1.44	10	16	53.92	23.20	5	41.9	13.71
NDUFS7	0.38	0.08	17	25	26.45	17.87	6	19.2	23.56
NDUFS8	0.19	0.06	27	33	22.37	11.84	9	42.4	20.66
NDUFV1	2.05	1.54	34	36	49.42	27.69	25	69.7	49.87
NDUFV2	2.37	1.65	15	26	23.19	17.66	12	54.2	27.39
NDUFV3	1.83	1.04	5	27	11.87	33.76	18	51.4	50.98

Complex I assembly factors									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
ACAD9	0.09	0.21	66	73	26.48	17.56	41	50.7	68.76
ATP5SL	0.83	1.21	2	8	13.62	11.64	6	24.1	29.27
COA1	-0.71	-0.67	5	12	42.62	16.06	7	54.8	16.69
ECSIT	0.17	0.27	22	37	13.01	11	15	41.8	49.15
FOXRED1	0.83	1.25	5	12	11.89	20.87	13	34.2	53.81
NDUFAF1	0.08	0.08	11	19	23.04	28.23	12	47.7	37.76
NDUFAF2	-0.42	-0.89	46	43	27.75	18.58	16	72.8	19.86
NDUFAF3	-0.32	-0.59	13	22	7.31	11.18	8	56.5	20.35
NDUFAF4	-0.38	-0.56	20	36	16.27	8.38	13	58.3	20.27

TIMMDC1	0.32	0.64	14	25	10.69	15.78	13	49.8	32.18
TMEM126B	0.52	0.58	4	9	30.73	34.57	3	51.8	18.89

Complex II subunits									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
SDHA	2.45	2.08	47	75	37.37	25.79	33	69.6	72.69
SDHB	2.54	2.00	34	47	31.98	27.86	8	68.2	31.63
SDHD	1.33	1.83	2	8	9.72	50.54	3	15.0	12.62

Complex III subunits									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
UQCC	-0.46	-0.90	6	14	6.48	27.17	7	24.4	34.60
UQCR10	0.51	-0.10	3	15	4.79	24.92	2	38.1	7.31
UQCRB	-0.32	-0.59	28	26	51.98	26.45	10	60.4	13.53
UQCRC1	-0.12	-0.19	106	115	28.13	30.15	24	61.5	52.65
UQCRC2	-0.12	-0.15	72	88	34.00	32.55	17	45.3	48.44
UQCRFS1	-0.10	-0.56	44	42	32.34	23.77	15	54.0	29.67
UQCRH	0.69	0.04	8	9	20.80	33.92	3	29.7	10.74
UQCRQ	0.07	-0.70	8	42	35.48	26.91	5	46.3	9.91

Complex IV subunits									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
COX4I1	0.46	0.66	60	86	18.62	14.71	17	56.2	19.58
COX5A	0.49	0.59	33	50	20.66	27.42	11	64.7	16.76
COX5B	0.30	0.76	31	25	48.75	17.81	8	45.7	13.70
COX6B1	0.83	1.50	15	11	35.92	24.95	5	73.3	10.19
COX6C	0.57	1.14	27	46	35.34	22.90	11	76.0	8.78
COX7A2	0.65	1.46	3	5	10.46	13.71	2	27.7	9.40
COX7A2L	0.26	0.01	9	25	42.38	36.60	6	49.1	12.62
COX7C	0.49	0.80	3	2	25.56	52.53	1	16.1	6.38
MT-CO2	0.59	1.11	33	31	30.30	19.09	7	32.2	25.57
MT-CYB	0.52	-0.30	3	5	53.37	23.74	2	5.0	42.72
NDUFA4	0.55	0.97	16	26	44.11	36.72	5	46.9	9.37

Complex V subunits									
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
MT-ATP8	0.25	0.01	2	8	13.62	11.64	3	48.5	7.99
MT-ATP6	-0.16	-0.31	2	6	11.60	5.92	1	4.4	24.82
ATP5SL	0.83	1.21	93	81	21.25	16.38	6	24.1	29.27
ATP5S	0.58	0.41	39	72	17.38	18.02	4	19.1	24.87
ATP5O	-0.04	-0.05	20	56	25.37	17.26	22	79.8	23.28
ATP5L	0.11	0.05	13	13	26.89	20.36	7	55.3	11.43
ATP5J2	0.28	0.11	13	21	20.47	14.73	3	52.1	10.92
ATP5J	0.16	0.00	22	55	29.83	27.13	6	51.9	12.59
ATP5I	0.23	0.00	48	88	25.99	8.77	6	66.7	7.93
ATP5H	-0.06	-0.04	73	101	21.01	16.31	21	87.0	18.49
ATP5F1	0.08	-0.03	8	10	18.42	18.96	22	58.6	28.91
ATP5E	-0.01	0.18	9	14	17.69	15.94	3	41.2	5.78
ATP5D	-0.02	0.10	44	64	15.66	12.45	5	29.2	17.49
ATP5C1	0.17	0.03	272	373	14.50	19.55	17	44.3	33.00
ATP5B	-0.02	0.00	2	3	17.79	10.57	21	75.8	56.56
ATP5A1	-0.03	-0.01	179	319	38.48	19.49	48	72.5	59.75

Mitoflavoproteins									
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
ACAD9	0.09	0.21	66	73	26.48	17.56	41	50.7	68.76
ACADM	-0.18	1.93	11	9	38.57	35.45	12	33.5	42.43
ACADVL	1.03	1.04	43	75	45.47	31.27	36	61.1	68.06
AIFM1	1.86	2.39	25	49	41.67	18.69	27	52.1	66.29
COQ6	-1.48	-1.17	9	11	21.6	28.52	7	20.3	48.59
CYB5R3	-0.1	-0.95	18	50	40.45	16.97	17	80.9	31.63
DHODH	0.64	0.96	31	24	17.69	16.83	18	57	42.87
DLD	3.18	3.03	7	9	35.08	20.08	15	43.6	54.18
ETFA	0.42	0.3	22	36	32.87	16.47	15	52.3	35.08
ETFDH	2.22	2.81	2	10	66.91	92.62	1	32.6	68.5
FOXRED1	0.83	1.25	5	12	11.89	20.87	13	34.2	53.81
GPD2	1.34	1.61	59	101	30.45	24.23	51	73.6	80.85
GSR	-0.27	0.54	2	2	56.8	52.89	5	15.5	47.27
IVD	0.39	0.95	10	8	16.25	11.62	8	18.7	46.32
NDUFV1	2.05	1.54	34	36	49.42	27.69	25	69.7	49.87
SDHA	2.45	2.08	47	75	37.37	25.79	33	69.6	72.69
SQRDL	3.54	5.11	9	5	52.09	36.08	27	60	49.96

Proteins with greater than a three-fold decrease abundance in conditioned RFK KD cells									
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**
SQRDL	3.54	5.11	9	5	52.09	36.08	27	60	49.96
DLD	3.18	3.03	7	9	35.08	20.08	15	43.6	54.18
NDUFA7	2.93	1.79	6	24	31.37	18.81	9	69	12.55
NDUFA12	2.68	1.59	6	24	15.78	19.89	10	82.8	17.11
FLRT3	2.57	2.55	3	4	20.38	95.24	16	27.7	73
SDHB	2.54	2.00	34	47	31.98	27.86	8	68.2	31.63
MMP1	2.48	2.68	16	3	23.27	57.72	12	33.9	54.01
AK2	2.45	2.11	5	13	45.71	17.45	12	65.5	25.61
SDHA	2.45	2.08	47	75	37.37	25.79	33	69.6	72.69
ENPP1	2.42	3.55	58	36	57.81	39.04	34	47.2	104.92
NDUFV2	2.37	1.65	15	26	23.19	17.66	12	54.2	27.39
NDUFS4	2.30	1.88	2	10	86.93	23.27	5	25.7	20.11
FDFT1	2.25	1.75	9	13	22.57	31	15	43.4	48.12
ETFDH	2.22	2.81	2	10	66.91	92.62	1	32.6	68.5
GJA1	2.17	1.93	6	6	24.58	26.4	6	24.6	43.01
APLP2	2.09	1.76	4	3	21.51	13.4	6	11.1	85.49
TSPAN13	1.94	1.68	3	3	82.02	28.7	3	21.1	22.15
AGPS	1.87	3.20	8	19	68.99	48.77	24	50	72.91
AIFM1	1.86	2.39	25	49	41.67	18.69	27	52.1	66.29
MMP14	1.86	1.71	13	6	52.27	11.23	11	18.7	65.89
MIPEP	1.79	2.36	4	4	47.15	29.49	9	16.8	80.64
QSOX2	1.79	1.94	3	7	23.81	24.1	11	20.3	77.53
PC	1.63	1.68	5	3	39.18	36.38	9	11.3	129.63
NDUFA2	1.60	1.74	6	11	61.2	38.33	5	42.4	10.92
TNS3	1.59	1.68	4	6	19.29	25.67	9	14.1	129.06

Appendix 7.4 SILAC data for pair 1: conditioned pLKO.1 143B cells and riboflavin rich 143B CyWT cells. Including proteins detected in one labelling direction.

Complex I subunits										
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
MT-ND1	2.08	-	2	-	7.75	-	3	8.8	35.66	59.46
MT-ND5	-	3.14	-	4	-	187.45	5	12.6	67.03	323.31
NDUFA10	1.62	1.69	3	7	6.48	8.71	13	38.3	40.75	143.49
NDUFA11	1.27	2.31	2	3	14.76	93.74	5	32.6	14.85	38.70
NDUFA12	2.64	2.52	4	8	44.32	160.62	7	57.2	17.11	61.74
NDUFA13	2.12	2.11	16	21	7.23	12.76	13	68.8	16.70	141.77
NDUFA2	2.73	2.90	3	7	0.85	3.12	6	56.6	10.92	81.66
NDUFA3	-1.48	1.44	2	2	244.52	14.86	4	50.0	9.28	21.64
NDUFA5	-	3.29	-	2	-	280.67	11	62.1	13.56	323.31
NDUFA6	-	4.53	-	4	-	113.18	4	33.1	17.87	37.95
NDUFA7	0.45	-	2	-	296.42	-	8	61.9	12.55	24.53
NDUFA8	0.79	2.02	5	3	215.30	95.76	5	36.0	20.11	70.44
NDUFA9	1.12	1.31	19	27	11.52	16.46	25	70.6	42.51	323.31
NDUFB1	0.73	0.62	3	3	41.14	59.58	3	34.5	6.96	9.62
NDUFB10	0.42	0.76	15	20	7.74	17.00	14	54.1	20.78	129.00
NDUFB11	-2.76	-	2	-	342.20	-	3	28.8	17.32	47.66
NDUFB2	1.68	1.70	2	2	5.11	2.34	2	8.6	12.06	3.56
NDUFB3	-	1.54	-	2	-	2.33	5	32.7	11.40	26.24
NDUFB4	1.98	1.88	9	16	30.03	11.33	6	45.0	11.40	63.92
NDUFB5	-1.78	3.71	3	3	166.43	161.37	5	35.0	14.03	72.82
NDUFB6	-4.15	2.20	3	3	176.41	107.71	8	60.2	15.67	40.57
NDUFB7	1.75	1.34	7	4	6.31	32.61	4	32.1	16.40	86.56
NDUFB8	1.82	-	7	-	12.35	-	7	41.3	18.48	40.77
NDUFB9	1.29	1.19	8	6	23.41	142.06	7	41.9	21.83	136.50
NDUFC2	1.06	1.13	6	8	189.37	4.74	10	51.3	14.19	34.08
NDUFS1	2.70	2.87	24	19	22.92	21.14	26	44.3	79.47	323.31
NDUFS2	0.70	0.72	21	28	6.88	27.83	22	65.9	52.55	323.31
NDUFS3	0.79	0.73	23	28	18.14	14.09	19	58.3	30.24	323.31
NDUFS5	-1.01	3.07	2	5	313.63	81.97	4	32.1	12.52	30.51
NDUFS6	2.57	5.69	2	3	11.71	157.95	5	39.5	13.71	170.93
NDUFS7	0.91	1.48	5	4	4.19	109.39	4	16.0	22.20	57.77
NDUFS8	0.96	2.10	5	4	101.28	148.45	9	28.1	23.71	126.52
NDUFV1	2.62	2.67	6	9	39.99	13.92	12	29.2	49.87	96.38
NDUFV2	2.24	2.61	5	7	94.45	82.03	10	37.8	27.39	97.48

Mitoflavoproteins										
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
ACAD8	-	1.52	-	2	-	7.51	3	7.5	18.49	45.07
ACAD9	-0.13	0.17	13	20	79.39	22.56	28	47.8	68.8	292.19
ACADM	1.41	2.2	7	7	19.45	15.41	11	38	46.6	217.05
ACADVL	0.89	1.72	15	17	11.85	18.59	30	49	68.1	323.31
AIFM1	3.27	3.49	6	10	34.84	25.81	25	52	66.9	323.31
COQ6	-0.79	-0.84	6	4	23.75	14.91	9	20.3	48.6	37.8
CYB5R3	-0.49	-0.69	13	20	8.07	21.99	14	56.5	31.6	323.31
DHODH	0.32	0.34	12	16	10.04	62.62	15	43.8	42.9	114.39
DLD	2.69	3.66	3	5	33.61	109.13	8	23.9	49.3	143.51
ETFA	1.63	1.82	2	3	3.27	11.97	12	60.9	30	188.97
ETFDH	2.6	2.67	4	3	44.12	33.73	10	20.3	68.5	99.38
FOXRED1	0.29	1.36	4	3	64.31	5.51	7	16.9	53.8	21.16
GCDH	2.67	2.4	2	3	76.14	27.72	3	15.1	48.1	24.47
GPD2	1.69	1.91	27	44	15.87	17.05	35	56.8	80.9	323.31
IVD	1.51	1.91	2	6	15.39	18.58	10	33.8	46.3	105.2
L2HGDH	0.56	0.72	6	10	9.09	59.74	13	37.1	50.3	132.98
NDUFV1	2.62	2.67	6	9	39.99	13.92	12	29.2	49.9	96.38
PPOX	1.52	2	2	3	9.35	71.15	7	21.2	50.8	12.72
SDHA	2.46	2.78	28	31	8.64	3.6	21	42.9	72.7	323.31
SQRDL	5.04	4.9	15	17	78.61	79.55	27	63.3	323.31	49.96
TXNRD2	-	-2.63	-	2	-	300.28	2	5.1	2.26	46.84

Proteins with greater than a two-fold decrease in abundance in conditioned pLKO.1 cells										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
SQRDL	5.04	4.90	15	17	78.61	79.55	27	63.3	49.96	323.31
ENPP1	4.01	2.74	15	39	68.43	61.70	27	35.2	104.92	323.31
AIFM1	3.27	3.49	6	10	34.84	25.81	25	52.0	66.90	323.31
MMP1	2.96	2.03	12	5	7.53	20.90	10	29.6	54.01	60.12
NDUFA2	2.73	2.90	3	7	0.85	3.12	6	56.6	10.92	81.66
NDUFS1	2.70	2.87	24	19	22.92	21.14	26	44.3	79.47	323.31
DLD	2.69	3.66	3	5	33.61	109.13	8	23.9	49.28	143.51
GCDH	2.67	2.40	2	3	76.14	27.72	3	15.1	48.13	24.47
NDUFA12	2.64	2.52	4	8	44.32	160.62	7	57.2	17.11	61.74
NDUFV1	2.62	2.67	6	9	39.99	13.92	12	29.2	49.87	96.38

Protein	Log_2 Forward Ratio H/L (normalised)	Log_2 inverted reverse ratio H/L (normalised)	Count (forward)	Count (reverse)	Ratio H/L variability [%] (forward)	Ratio H/L variability [%] (reverse)	Unique peptides	Sequence coverage [%]	Molecular weight [kDa]	Score
ETFDH	2.60	2.67	4	3	44.12	33.73	10	20.3	68.50	99.38
AK2	2.59	2.77	2	5	161.33	161.05	10	60.3	25.61	89.55
NDUFS6	2.57	5.69	2	3	11.71	157.95	5	39.5	13.71	170.93
MMP14	2.51	1.18	9	6	16.53	91.75	9	18.6	65.89	44.67
COL12A1	2.46	1.60	25	18	36.02	110.77	43	21.2	324.57	89.55
SDHA	2.46	2.78	28	31	8.64	3.60	21	42.9	72.69	323.31
FDFT1	2.43	1.56	12	9	29.13	12.85	14	40.3	48.12	228.33
IDH2	2.28	3.18	4	8	43.43	58.62	19	38.9	50.91	233.38
NDUFV2	2.24	2.61	5	7	94.45	82.03	10	37.8	27.39	97.48
SLC25A1	2.20	2.04	6	13	17.76	53.71	10	38.3	34.01	96.00
JUP	2.20	2.95	5	4	251.64	56.59	16	27.1	81.74	48.06
NDUFA13	2.12	2.11	16	21	7.23	12.76	13	68.8	16.70	141.77
PRDX3	2.11	2.11	9	10	33.07	3.44	11	36.1	25.84	114.55
NT5E	2.06	1.21	7	7	32.08	8.29	13	35.9	63.37	204.55
TST	2.03	2.49	2	6	8.47	86.71	6	34.3	33.43	142.29
SDHB	2.01	2.82	12	13	29.79	9.53	14	40.7	31.63	78.96
PXDN	2.01	1.02	9	7	27.06	40.16	13	12.9	165.27	67.07
ALDH2	2.00	1.97	4	5	9.14	35.31	12	27.9	56.38	86.83
AGPS	1.98	2.77	10	10	32.45	7.26	16	36.5	72.91	232.92
NDUFB4	1.98	1.88	9	16	30.03	11.33	6	45.0	14.03	63.92
SLC4A7	1.98	2.32	6	8	113.79	12.89	19	22.0	127.36	131.12
MT-CO2	1.96	2.30	12	7	7.20	12.19	7	32.2	25.57	119.53
CDK17	1.93	1.07	2	3	25.49	23.85	4	10.5	59.53	24.40
P56181-2	1.88	1.94	2	3	25.07	33.41	11	36.2	50.98	302.43
ECH1	1.87	2.04	2	3	31.33	116.40	6	34.1	35.82	82.96
NIPSNAP1	1.87	1.89	4	7	10.47	81.95	8	28.5	33.31	118.57
IARS2	1.85	1.88	8	10	37.65	32.16	25	30.6	113.79	153.52
SCO2	1.83	1.11	4	5	53.59	40.20	7	33.8	29.81	124.50
AHR	1.83	1.85	4	5	77.61	61.76	13	20.2	96.15	82.88
SDHC	1.82	1.49	2	3	7.36	49.18	3	42.2	12.47	12.68
MCCC2	1.82	2.03	6	13	16.64	37.07	18	42.6	61.33	127.30
AIM1	1.78	2.58	5	7	23.97	73.80	26	24.0	188.67	132.19
OAT	1.77	2.28	15	12	23.08	17.72	15	49.0	48.53	300.01
COLEC12	1.75	2.66	2	3	64.97	24.70	4	6.5	81.51	11.36
NDUFB7	1.75	1.34	7	4	6.31	32.61	4	32.1	16.40	86.56
MCCC1	1.73	1.73	2	3	41.76	29.78	4	10.8	80.47	83.79
PC	1.71	2.23	4	12	110.58	76.02	21	26.2	129.63	77.22
OXCT1	1.71	1.99	4	5	18.64	11.43	8	31.0	56.16	323.31
ME2	1.71	2.24	5	9	13.06	115.54	10	26.2	65.44	151.55
AK3	1.70	2.11	5	5	4.45	35.08	9	43.2	25.57	120.92

Protein	Log_2 Forward Ratio H/L (normalised)	Log_2 inverted reverse ratio H/L (normalised)	Count (forward)	Count (reverse)	Ratio H/L variability [%] (forward)	Ratio H/L variability [%] (reverse)	Unique peptides	Sequence coverage [%]	Molecular weight [kDa]	Score
ALDH7A1	1.69	2.45	2	3	39.10	38.42	9	28.6	55.37	50.20
GPD2	1.69	1.91	27	44	15.87	17.05	35	56.8	80.85	323.31
NDUFB2	1.68	1.70	2	2	5.11	2.34	2	8.6	12.06	3.56
SNTB1	1.68	1.57	5	5	34.95	59.20	10	29.4	58.06	46.24
SCD	1.67	1.54	2	5	6.03	88.29	7	27.9	41.52	74.31
GFM1	1.66	2.12	5	6	23.06	11.35	17	26.1	83.47	111.65
LAMC1	1.64	1.53	7	7	14.55	16.36	17	15.2	177.60	148.41
HIP1	1.64	1.13	2	4	78.43	40.08	6	9.4	110.66	20.96
ITGB5	1.63	1.75	6	9	48.54	85.43	11	16.4	88.05	51.08
ETFA	1.63	1.82	2	3	3.27	11.97	12	60.9	30.03	188.97
VPS13C	1.62	1.91	27	20	23.65	36.97	63	23.9	403.08	189.54
NDUFA10	1.62	1.69	3	7	6.48	8.71	13	38.3	40.75	143.49
ACO2	1.58	1.92	6	6	8.57	13.02	21	38.8	85.42	230.94
HINT2	1.56	1.86	2	5	1.98	28.28	5	47.2	17.16	115.70
SVIL	1.54	1.77	14	21	24.97	40.87	54	41.6	200.84	323.31
PDP1	1.53	1.87	5	7	20.70	12.07	12	31.3	61.05	132.11
PCYOX1	1.53	1.46	3	3	12.47	13.21	9	22.2	56.64	22.31
PPOX	1.52	2.00	2	3	9.35	71.15	7	21.2	50.77	12.72
IVD	1.51	1.91	2	6	15.39	18.58	10	33.8	46.32	105.20
CLPP	1.51	1.82	2	3	3.52	127.88	9	33.6	30.18	133.64
NADK2	1.50	1.77	3	5	42.41	135.03	8	24.4	45.97	76.94
FASN	1.50	1.42	24	35	48.74	40.33	51	28.2	273.42	323.31
ECHS1	1.49	1.37	13	18	3.36	5.81	15	36.6	31.39	181.93
IDH3G	1.47	2.50	2	4	5.82	65.15	6	15.0	42.79	21.70
ACAA2	1.46	1.79	6	7	9.62	11.53	12	45.8	41.92	270.58
SPRYD4	1.46	1.41	3	3	82.45	29.20	8	42.5	23.13	18.18
ETFB	1.45	1.49	11	10	23.97	1.16	16	46.3	27.84	49.09
KRAS	1.45	1.13	19	21	40.00	20.76	2	47.6	21.66	323.31
GLUD1	1.44	1.65	15	15	19.27	8.90	17	37.6	61.40	238.94
PRDX5	1.44	1.58	4	2	20.14	7.32	12	54.3	17.03	172.74
HIBCH	1.41	1.86	3	3	6.78	17.91	9	26.2	43.48	43.73
ACADM	1.41	2.20	7	7	19.45	15.41	11	38.0	46.59	217.05
SLC25A3	1.33	1.61	36	46	22.74	9.98	19	42.9	39.96	259.69
CS	1.32	1.77	10	13	24.97	13.87	17	51.9	51.71	323.31
NCKAP1	1.32	1.17	3	4	6.14	26.42	9	9.0	128.79	57.48
GOT2	1.31	1.41	16	20	20.18	41.24	22	45.8	47.52	323.31
AK4	1.30	1.83	6	3	14.13	19.41	7	33.2	25.27	39.70
NDUFB9	1.29	1.19	8	6	23.41	142.06	7	41.9	21.83	136.50
CYFIP1	1.28	1.11	9	17	38.27	80.95	11	21.2	145.18	74.85
GBAS	1.28	1.29	4	6	26.81	37.12	5	23.8	33.74	41.83

Protein	Log_2 Forward Ratio H/L (normalised)	Log_2 inverted reverse ratio H/L (normalised)	Count (forward)	Count (reverse)	Ratio H/L variability [%] (forward)	Ratio H/L variability [%] (reverse)	Unique peptides	Sequence coverage [%]	Molecular weight [kDa]	Score
NDUFA11	1.27	2.31	2	3	14.76	93.74	5	32.6	14.85	38.70
ITGA1	1.26	1.85	4	8	41.15	17.18	13	13.6	130.85	42.00
CPT2	1.26	1.52	4	10	31.28	117.33	13	27.7	73.78	111.15
GNB4	1.26	1.10	3	4	14.92	11.57	2	11.5	37.57	3.79
LDLR	1.25	1.48	10	12	44.22	6.33	12	22.6	75.85	110.01
CD99L2	1.22	1.69	5	3	14.59	46.41	2	10.6	20.03	10.92
CD109	1.19	1.44	2	13	4.72	43.47	20	21.5	161.69	98.98
HELZ2	1.18	1.59	4	6	159.36	125.89	16	7.4	294.65	27.15
ICAM1	1.18	2.11	2	2	15.63	24.15	5	13.5	57.83	31.37
IDH3A	1.18	1.72	6	12	20.94	30.81	9	31.4	39.59	83.68
SARS2	1.15	1.67	3	6	111.48	35.01	9	26.6	58.28	79.79
KIAA0391	1.15	1.38	2	2	19.62	40.77	6	14.1	56.49	26.65
DNAJC10	1.15	1.51	5	6	9.75	17.67	12	19.2	91.08	251.30
ACACA	1.13	1.09	3	6	131.42	25.69	14	9.9	257.24	54.98
NDUFA9	1.12	1.31	19	27	11.52	16.46	25	70.6	42.51	323.31
ARHGEF40	1.12	1.19	6	9	4.79	24.00	15	12.9	164.66	96.91
IDH3B	1.11	1.36	7	8	18.79	7.57	13	39.5	42.18	54.63
HSD17B10	1.11	1.36	8	9	11.26	8.31	11	63.6	26.92	311.66
PIK3C2A	1.11	1.05	8	9	18.32	30.79	30	21.0	190.68	99.79
NDUFC2	1.06	1.13	6	8	189.37	4.74	10	51.3	14.19	34.08
PITRM1	1.06	1.19	7	12	22.47	8.50	22	30.6	117.41	323.31
PDPR	1.06	1.74	3	4	28.73	18.91	8	11.6	99.36	25.84
HIBADH	1.05	1.52	3	3	46.60	7.34	10	40.5	35.33	129.53
GLS	1.05	1.25	6	6	15.42	18.10	3	39.8	65.46	323.31
ACOT1	1.03	1.56	4	7	26.38	31.31	2	36.8	46.28	56.57
RTCB	1.01	1.05	13	11	6.18	7.03	22	50.1	55.21	255.26
ACAT1	1.01	1.26	6	5	19.68	21.32	12	39.1	45.20	168.02
ALDH1B1	1.00	1.11	7	8	39.60	55.31	14	38.5	57.21	144.32

Appendix 7.5 SILAC data for pair 2: conditioned RFK KD 143B cells and conditioned pLKO.1 143B cells. Including proteins detected in one labelling direction.

Complex I subunits										
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
MT-ND1	-0.15	-	2	-	119.52	-	2	6.3	35.66	43.75
MT-ND2	-0.51	-	2	-	9.54	-	1	2.6	38.96	2.16
MT-ND4	-0.50	-	4	-	7.48	-	4	10.2	51.58	4.42
MT-ND5	-0.66	-	7	-	7.58	-	5	12.6	67.03	156.08
NDUFA1	-0.95	-	2	-	11.76	-	1	18.6	8.07	2.44
NDUFA10	-0.77	-	18	-	28.15	-	9	23.9	40.75	103.73
NDUFA11	-1.10	-	9	-	24.20	-	6	32.6	14.85	27.85
NDUFA12	1.18	-	4	-	109.88	-	3	29.0	17.11	28.09
NDUFA13	-1.12	-	28	-	6.27	-	11	63.9	16.70	87.50
NDUFA2	-0.41	-	5	-	1.55	-	3	32.3	10.92	14.82
NDUFA3	-0.52	-	6	-	10.13	-	4	50.0	9.28	51.49
NDUFA5	-0.57	-0.60	17	4	4.24	0.92	11	64.7	13.56	323.31
NDUFA6	-0.41	-	10	-	24.62	-	5	39.6	17.87	19.74
NDUFA7	-0.41	-	6	-	7.01	-	5	49.6	12.55	12.56
NDUFA8	-1.05	-	8	-	24.46	-	5	39.0	20.11	29.61
NDUFA9	-0.25	-0.03	30	5	8.64	4.62	17	56.2	42.51	98.16
NDUFAB1	-0.51	-	5	-	7.21	-	3	11.5	17.42	23.97
NDUFB1	-0.20	-	5	-	10.85	-	3	34.5	6.96	14.32
NDUFB10	-0.31	-0.08	24	7	7.92	5.76	11	48.8	20.78	111.69
NDUFB11	-0.16	-0.10	8	2	6.34	3.18	3	34.6	17.32	57.48
NDUFB2	-0.87	-	2	-	8.29	-	2	8.6	12.06	3.26
NDUFB3	-0.62	-	5	-	1.87	-	4	29.6	11.40	23.87
NDUFB4	-0.93	-	13	-	2.79	-	6	48.1	15.21	74.57
NDUFB5	-0.12	-0.13	17	2	2.87	0.27	8	29.6	21.75	121.64
NDUFB6	-0.05	0.14	14	2	25.97	6.54	3	43.8	15.49	46.91
NDUFB7	-0.89	-	5	-	57.42	-	5	32.8	16.40	57.64
NDUFB8	-0.92	-	8	-	4.61	-	5	36.1	18.48	36.04
NDUFB9	-0.62	-	7	-	6.06	-	6	41.9	21.83	63.43
NDUFC1	-	-	-	-	-	-	1	14.5	8.73	2.59
NDUFC2	-0.51	-0.18	17	2	7.84	2.96	11	60.5	14.19	24.47
NDUFS1	-0.65	-0.04	26	9	20.44	7.02	20	37.3	79.47	212.73
NDUFS2	-0.50	-0.61	48	11	10.39	11.72	23	67.0	52.55	323.31
NDUFS3	-0.58	-0.61	40	9	5.31	4.25	16	52.3	30.24	309.72
NDUFS4	-0.37	-	3	-	8.53	-	3	20.0	20.11	4.37
NDUFS5	-1.06	-	8	-	24.76	-	6	39.6	12.52	12.47

NDUFS6	-0.44	-	9	-	11.89	-	4	39.5	13.71	27.99
NDUFS7	-0.64	-	14	-	8.15	-	4	16.0	22.20	75.12
NDUFS8	-0.50	-0.62	9	3	5.39	1.77	5	23.8	23.71	130.22
NDUFV1	-0.36	-0.20	11	3	19.60	145.76	10	24.8	49.87	24.10
NDUFV2	-0.66	-	5	-	16.65	-	4	21.3	27.39	70.19

Mitoflavoproteins										
Protein	**Log_2 Forward ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Score**	**Molecular weight [kDa]**
ACAD9	0.03	-0.20	49	15	16.22	4.10	32	43.3	68.76	257.83
ACADM	-0.07	-	10	-	13.64	-	9	25.9	46.59	165.95
ACADVL	-0.23	-0.27	58	14	25.42	12.91	24	44.4	68.06	305.61
AIFM1	-0.99	-	13	-	16.71	-	12	29.6	66.29	103.32
COQ6	-0.05	-0.71	7	2	19.13	11.11	5	14.9	48.59	30.61
CYB5R3	0.09	0.01	35	5	8.31	9.88	17	59.4	31.63	323.31
DHODH	0.12	-0.14	21	5	13.43	11.09	16	51.1	42.87	128.63
ETFA	-0.79	-0.36	18	3	7.59	1.11	11	53.9	30.03	180.41
FOXRED1	-0.57	-	6	-	45.17	-	6	15.3	52.10	11.58
GCDH	-0.36	-	2	-	28.19	-	2	9.8	47.36	18.21
GPD2	-0.25	-0.21	37	14	16.62	18.40	27	47.9	80.85	254.71
GSR	-0.23	-	7	-	7.94	-	5	17.5	47.27	17.55
IVD	-0.77	-	6	-	18.78	-	7	27.2	42.72	20.14
L2HGDH	-0.29	-	10	-	15.23	-	9	26.8	50.32	56.28
NDUFV1	-0.36	-0.20	11	3	19.60	145.76	10	24.8	49.87	24.10
SDHA	-0.68	0.05	19	9	23.58	7.70	13	29.4	72.69	145.76

Appendix 7.6 SILAC data for conditioned SLC25A32 KO cells and conditioned HAP1 cells

Complex I subunits										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
MT-ND1	-0.52	0.89	2	3	5.40	29.34	3	8.8	35.66	3.98
NDUFA10	-0.65	-0.36	12	7	7.08	9.42	10	30.7	40.75	87.86
NDUFA11	0.29	-0.99	8	5	12.95	26.53	7	44.7	14.85	27.09
NDUFA13	0.39	0.84	20	10	7.55	8.88	9	63.9	16.70	43.66
NDUFA3	0.13	1.49	6	4	20.35	23.90	4	50	9.28	16.98
NDUFA5	-0.03	0.53	8	4	13.84	26.07	8	73.3	13.46	31.86
NDUFA6	-0.61	0.79	6	3	10.45	4.31	5	37.7	17.87	17.79
NDUFA7	-0.30	-1.06	3	3	14.66	12.69	3	24.8	12.55	9.84

NDUFA8	0.37	0.70	11	10	10.11	40.03	10	64.5	20.11	67.15
NDUFA9	0.31	0.61	35	17	8.77	21.23	21	61	42.51	147.26
NDUFB10	0.18	-0.32	43	39	9.71	30.01	22	75.6	20.78	165.38
NDUFB11	0.14	0.46	16	9	4.89	30.15	8	51.6	17.32	91.53
NDUFB3	-0.18	-0.77	11	5	6.76	26.43	6	32.7	11.40	22.66
NDUFB4	-0.60	-0.64	11	6	16.58	36.99	7	49.2	14.03	37.38
NDUFB5	0.03	-0.46	18	12	12.38	54.30	7	29.6	21.75	84.69
NDUFB6	0.37	0.06	20	14	17.31	13.51	10	64.1	15.49	64.92
NDUFB7	-0.74	-0.11	10	5	17.64	21.69	6	48.2	16.40	26.94
NDUFB9	-0.86	-0.07	13	8	5.14	32.57	11	69.3	21.83	95.96
NDUFC2	0.44	-0.21	10	8	12.41	30.97	9	52.9	14.19	19.21
NDUFS1	-0.76	-0.90	8	14	58.06	8.33	16	36.1	73.53	80.69
NDUFS2	-0.12	0.74	26	17	6.68	11.82	24	71.7	52.55	263.44
NDUFS3	-0.12	0.70	22	21	9.49	26.41	17	57.6	30.24	152.38
NDUFS5	-0.77	-0.87	6	2	9.06	27.07	5	46.2	12.52	13.63
NDUFS6	-0.90	-0.43	9	5	12.52	42.59	6	61.3	13.71	18.40
NDUFS7	-0.03	0.86	7	4	12.65	21.14	6	31.5	23.56	28.85
NDUFS8	0.04	0.71	16	5	9.08	7.42	7	44.8	23.705	42.664
NDUFV1	-0.79	-1.15	27	8	15.00	102.65	18	58.5	49.87	109.29
NDUFV2	-0.74	-1.03	15	9	9.50	48.44	10	43	27.39	65.96

Complex I assembly factors										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
ACAD9	-0.43	-0.07	38	30	15.28	12.32	27	46.4	68.76	158.80
COA1	-0.12	-0.40	9	4	13.75	9.59	6	54.8	16.69	24.53
ECSIT	-0.36	0.03	14	10	23.53	7.15	13	43.9	49.15	173.54
FOXRED1	-0.78	0.14	6	6	46.18	128.55	8	24.7	53.81	111.29
NDUFAF2	0.23	0.73	21	24	8.51	18.21	16	74.6	19.86	207.05
NDUFAF3	0.24	1.23	12	6	6.24	4.52	7	44	20.35	98.76
NDUFAF4	0.23	1.27	19	9	9.03	18.47	14	56.6	20.27	57.02
NDUFAF6	-0.48	0.01	5	2	11.80	42.90	7	34.5	32.85	26.96
NDUFAF7	0.06	-0.59	3	4	30.01	41.13	5	13.8	49.24	12.84
TIMMDC1	0.12	0.33	20	18	10.56	8.56	16	54.7	32.18	185.50
TMEM70	0.10	1.31	12	11	5.44	6.41	6	25.8	28.97	57.09

Complex II subunits										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
SDHA	-0.87	-1.74	19	22	8.73	109.95	17	46.4	67.30	256.46
SDHB	-1.57	-2.33	3	4	39.72	39.96	6	19.6	31.63	12.28

Complex III subunits										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
CYC1	-1.03	-1.02	73	53	7.63	14.76	15	56.9	35.42	323.31
UQCR10	-0.54	-1.87	4	3	15.60	84.12	5	57.1	7.31	25.36
UQCRB	-1.51	-1.44	24	21	6.38	38.34	14	71.2	13.53	70.41
UQCRC1	-1.28	-1.00	47	27	8.81	14.80	21	57.1	52.65	323.31
UQCRC2	-1.42	-1.10	28	32	10.24	11.02	18	65.3	48.44	323.31
UQCRFS1	-0.95	-0.79	18	24	17.58	21.59	11	48.9	29.67	202.15
UQCRH	-1.00	-1.37	7	5	7.40	31.76	6	57.1	10.74	22.95
UQCRQ	-1.44	-1.09	9	5	7.42	40.93	6	48.8	9.91	27.38

Complex IV subunits										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
COX7A2L	-1.16	-0.50	7	3	19.39	19.72	6	67.5	12.62	38.14
COX6C	-0.11	-0.28	26	12	9.93	27.56	8	52	8.78	26.72
COX6B1	-0.88	-1.60	3	2	10.37	47.58	4	59.3	10.19	13.23
COX5B	-1.19	-0.95	11	16	13.23	29.82	9	65.1	13.70	58.64
COX5A	0.09	-0.34	25	18	14.14	28.78	10	57.3	16.76	88.56
COX4I1	0.26	-0.29	82	44	10.80	55.34	22	65.7	19.58	189.37
COX17	2.55	2.86	3	4	44.81	25.05	1	25.4	6.92	16.41
COX11	-0.14	-0.15	5	3	11.11	45.36	6	38.8	31.43	30.86

Complex V subunits										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
ATP5A1	-0.14	-0.02	315	295	11.10	14.87	1	71.6	59.75	323.31
ATP5B	-0.20	0.06	325	246	6.12	9.69	34	75.2	56.56	323.31
ATP5C1	-0.16	0.03	59	42	4.56	9.16	1	46	33.00	235.78
ATP5D	-0.21	0.10	26	14	9.81	79.47	6	35.1	17.49	45.61
ATP5E	-0.13	-0.88	6	10	4.28	47.60	7	70.6	5.78	17.06
ATP5F1	-0.29	-0.05	88	88	11.38	5.24	23	59.4	28.91	323.31
ATP5H	-0.37	0.13	50	53	5.85	36.33	24	96.3	18.49	234.20
ATP5I	-0.17	-0.09	17	13	6.28	27.87	9	71	7.93	41.64
ATP5J	-0.26	-0.46	19	11	8.74	33.41	13	66.7	12.59	143.37
ATP5J2	-0.39	-0.38	26	23	11.41	31.23	2	53.4	10.36	7.36
ATP5J2	-0.31	-0.68	16	15	12.81	24.64	3	56.4	10.92	97.40
ATP5O	-0.26	-0.03	77	70	5.43	22.32	26	79.8	23.28	323.31
ATP5S	-0.31	0.12	6	6	25.58	42.44	6	35.8	24.87	18.54
MT-ATP6	-1.56	-0.67	5	4	27.40	131.82	2	6.2	24.82	10.60
MT-ATP8	-1.26	-1.24	4	2	14.98	20.48	5	55.9	7.99	25.95

Mitoflavoproteins										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
ACAD8	-2.53	0.82	6	2	89.17	92.34	6	14.0	45.07	15.15
ACAD9	-0.43	-0.07	38	30	15.28	12.32	27	46.4	68.76	158.80
ACADM	-0.37	-1.47	16	12	10.23	18.12	14	38.5	46.59	84.37
ACADSB	-0.57	-0.57	22	12	11.54	11.43	20	62.7	47.49	127.54
ACADVL	-0.95	-1.57	23	23	19.59	33.42	29	51.5	68.06	238.30
AIFM1	1.05	-0.52	57	43	12.04	18.88	32	57.4	66.90	323.31
COQ6	-1.25	0.58	13	7	18.71	9.31	12	43.6	48.59	51.46
CYB5R3	-0.28	-0.27	59	49	7.23	14.48	21	81.3	31.63	323.31
DHODH	1.16	0.36	9	9	14.70	5.64	16	47.6	42.87	69.43
DLD	-0.23	-1.01	12	6	11.52	28.15	9	29.7	49.28	95.99
ETFA	-1.62	-0.51	50	27	6.57	8.93	20	69.4	35.08	218.94
ETFDH	-0.72	-1.08	2	7	23.10	77.34	9	18.5	68.50	20.91
FDXR	-2.38	-1.30	23	12	9.88	51.71	23	61.1	53.84	153.88
FOXRED1	-0.78	0.14	6	6	46.18	128.55	8	24.7	53.81	111.29
GCDH	-0.90	-0.66	6	4	22.01	85.87	7	27.9	48.13	89.81
GPD2	1.14	-0.02	55	51	23.07	31.53	44	71.5	80.85	277.12
GSR	0.03	0.24	9	5	10.35	25.13	9	26.9	51.70	30.47

IVD	-1.42	-2.79	17	4	25.92	111.80	13	38.8	46.32	86.35
L2HGDH	-1.85	-0.42	15	6	15.99	93.72	17	44.7	50.32	77.49
MAOA	1.86	-0.40	22	18	42.99	26.97	20	51.6	59.68	85.66
NDUFV1	-0.79	-1.15	27	8	15.00	102.65	18	58.5	49.87	109.29
PPOX	-1.00	-1.29	3	4	14.67	131.36	8	25.6	50.77	22.70
PRODH	-1.56	-2.45	28	22	20.17	12.20	26	56.3	68.00	258.29
SDHA	-0.87	-1.74	19	22	8.73	109.95	17	46.4	67.30	256.46
TXNRD2	-1.31	-1.14	9	2	37.31	6.77	10	37.3	53.97	99.02

Folate metabolising enzymes										
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**
MTHFD1L	-0.40	0.61	31	17	53.49	11.70	28	39.7	105.79	136.15
MTHFD2	-0.26	-0.21	31	5	8.82	43.94	13	64.9	37.90	192.14
SHMT2	0.01	-0.36	67	34	11.30	8.18	27	71.2	53.45	323.31

Proteins with greater than a two-fold decrease in abundance in SLCK25A32 KO cells											
Protein	**Log_2 Forward Ratio H/L (normalised)**	**Log_2 inverted reverse ratio H/L (normalised)**	**Count (forward)**	**Count (reverse)**	**Ratio H/L variability [%] (forward)**	**Ratio H/L variability [%] (reverse)**	**Unique peptides**	**Sequence coverage [%]**	**Molecular weight [kDa]**	**Score**	**Mitochondrial localisation?**
FAM213A	-2.62	-3.44	8	2	30.46	67.42	7	36.7	24.46	27.22	*Non-mitochondrial*
SLC52A2	-2.50	-1.74	7	7	35.71	39.35	3	4.5	45.78	10.19	*Non-mitochondrial*
DGAT1	-2.42	-2.07	14	7	9.67	64.14	14	39.3	55.28	133.75	*Non-mitochondrial*
FDXR	-2.38	-1.30	23	12	9.88	51.71	23	61.1	53.84	153.88	*Mitochondrial*
BDH1	-2.14	-1.21	15	19	11.11	40.37	17	57.7	38.16	96.58	*Mitochondrial*
RFTN1	-1.90	-1.27	4	6	44.84	131.33	10	30.6	63.15	34.93	*Non-mitochondrial*
AASS	-1.79	-1.30	30	26	32.49	40.52	34	45.8	102.13	215.80	*Mitochondrial*
WWOX	-1.78	-1.37	5	4	27.33	138.72	12	31.2	46.68	36.34	*Non-mitochondrial*
USP7	-1.71	-1.77	43	27	19.39	110.94	45	47.6	126.27	279.38	*Non-mitochondrial*
EEF1D	-1.66	-1.18	37	12	20.63	11.08	2	79.4	31.12	323.31	*Non-mitochondrial*
BPHL	-1.64	-1.24	15	6	21.93	170.15	15	56.2	31.11	68.11	*Predicted mitochondrial*
IDH2	-1.58	-1.46	37	27	15.85	45.19	26	51.1	50.91	323.31	*Mitochondrial*
ALDH6A1	-1.58	-1.40	24	19	76.37	15.19	20	44.6	56.23	175.28	*Non-mitochondrial*
SDHB	-1.57	-2.33	3	4	39.72	39.96	6	19.6	31.63	12.28	*Mitochondrial*
PRODH	-1.56	-2.45	28	22	20.17	12.20	26	56.3	68.00	258.29	*Mitochondrial*
ATP2B4	-1.51	-2.55	21	19	13.06	48.91	20	35.3	132.59	179.93	*Non-mitochondrial*
UQCRB	-1.51	-1.44	24	21	6.38	38.34	14	71.2	13.53	70.41	*Mitochondrial*
UQCRQ	-1.44	-1.09	9	5	7.42	40.93	6	48.8	9.91	27.38	*Mitochondrial*
STX1B	-1.44	-2.24	2	2	59.54	20.19	5	26.4	31.79	15.72	*Non-mitochondrial*

IVD	-1.42	-2.79	17	4	25.92	111.80	13	38.8	46.32	86.35	*Mitochondrial*
UQCRC2	-1.42	-1.10	28	32	10.24	11.02	18	65.3	48.44	323.31	*Mitochondrial*
SLC12A7	-1.42	-1.00	2	2	43.57	29.75	6	11.4	119.10	21.64	*Non-mitochondrial*
MACROD1	-1.41	-1.40	6	7	6.91	32.59	13	53.2	35.51	55.19	*Non-mitochondrial*
CA5B	-1.36	-1.04	7	2	9.75	108.54	9	31.5	36.43	42.57	*Mitochondrial*
TXNRD2	-1.31	-1.14	9	2	37.31	6.77	10	37.3	53.97	99.02	*Mitochondrial*
ACAA2	-1.28	-1.02	49	36	9.00	44.65	26	80.4	41.92	323.31	*Mitochondrial*
UQCRC1	-1.28	-1.00	47	27	8.81	14.80	21	57.1	52.65	323.31	*Mitochondrial*
MT-ATP8	-1.26	-1.24	4	2	14.98	20.48	5	55.9	7.99	25.95	*Mitochondrial*
SLC25A32	-1.25	-2.30	4	2	105.34	1.41	7	29.2	35.41	17.66	*Mitochondrial*
GNAZ	-1.25	-1.21	4	5	31.10	31.16	8	33.0	40.92	30.80	*Non-mitochondrial*
FECH	-1.17	-1.47	18	6	16.82	44.90	16	49.4	47.86	100.46	*Mitochondrial*
WLS	-1.15	-1.12	13	5	7.55	24.20	11	20.1	62.25	35.12	*Non-mitochondrial*
CD200	-1.11	-1.09	4	4	29.72	43.35	4	21.2	30.33	32.62	*Non-mitochondrial*
CYP2S1	-1.05	-1.49	10	4	17.40	37.38	12	33.7	55.82	80.00	*Non-mitochondrial*
ANK2	-1.04	-1.33	5	20	39.74	50.58	19	15.1	204.75	39.28	*Non-mitochondrial*
CYC1	-1.03	-1.02	73	53	7.63	14.76	15	56.9	35.42	323.31	*Mitochondrial*
DHRS1	-1.03	-1.15	10	8	11.12	72.88	15	60.4	33.91	108.72	*Predicted mitochondrial*
UQCRH	-1.00	-1.37	7	5	7.40	31.76	6	57.1	10.74	22.95	*Mitochondrial*

Proteins with greater than a two-fold increase in abundance in SLCK25A32 KO cells

Protein	Log_2 Forward Ratio H/L (normalised)	Log_2 inverted reverse ratio H/L (normalised)	Count (forward)	Count (reverse)	Ratio H/L variability [%] (forward)	Ratio H/L variability [%] (reverse)	Unique peptides	Sequence coverage [%]	Molecular weight [kDa]	Score	Mitochondrial localisation?
MAGEB2	5.42	5.94	3	3	101.05	44.43	6	22.9	35.28	42.49	*Non-mitochondrial*
RAB34	3.69	3.37	7	5	115.19	212.05	12	47.9	29.04	32.88	*Non-mitochondrial*
SSX2	3.29	5.01	8	5	139.23	83.76	10	53.7	21.62	42.46	*Non-mitochondrial*
ABCB6	2.75	3.71	5	6	130.43	120.31	13	26.4	88.60	103.33	*Mitochondrial*
SLC7A3	2.61	2.14	9	6	84.63	19.75	5	11.5	67.17	121.63	*Non-mitochondrial*
COX17	2.55	2.86	3	4	44.81	25.05	1	25.4	6.92	16.41	*Mitochondrial*
NPR2	2.23	2.22	3	4	128.93	131.62	9	12.1	117.02	18.51	*Non-mitochondrial*
SLC43A3	2.20	3.55	4	5	61.30	112.56	4	10.8	54.53	8.60	*Non-mitochondrial*
ABCB1	1.90	1.00	26	6	24.98	20.85	26	27.7	134.03	207.37	*Non-mitochondrial*
HMOX1	1.78	1.80	37	24	11.40	20.12	15	59.7	32.82	186.54	*Non-mitochondrial*
ABCD1	1.43	1.65	3	2	35.61	9.62	3	6.4	82.94	14.88	*Predicted mitochondrial*
VLDLR	1.43	1.63	19	7	16.04	9.38	22	34.0	93.38	150.93	*Non-mitochondrial*
IGSF8	1.39	1.14	3	4	15.79	31.20	5	12.1	65.03	30.77	*Non-mitochondrial*
PROCR	1.37	2.74	3	2	10.87	13.35	2	10.9	26.67	10.54	*Non-mitochondrial*
AKAP12	1.34	1.64	96	92	11.07	31.84	70	58.6	180.99	323.31	*Non-mitochondrial*
NES	1.24	2.42	71	124	23.47	35.79	57	44.5	177.44	323.31	*Non-mitochondrial*
BST2	1.19	2.06	12	13	13.55	6.38	4	19.6	18.39	21.08	*Non-mitochondrial*

METTL7A	1.16	1.87	21	13	8.82	23.15	9	38.5	28.32	104.88	*Non-mitochondrial*
SCD5	1.13	1.37	3	2	48.50	10.98	3	17.6	28.98	12.24	*Non-mitochondrial*
VIM	1.10	1.47	109	85	20.13	19.98	32	65.9	53.65	323.31	*Non-mitochondrial*
GNB4	1.02	1.20	15	12	11.08	8.44	10	67.4	37.57	236.88	*Non-mitochondrial*

Appendix 7.7 Transcriptome data for conditioned SLC25A32 KO cells and conditioned HAP1 cells

Complex I subunits													
								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
MT-ND1	14595.92	0.70	0.31	2.27	2.32E-02	1.02E-01	MT	948.1	728.1	460.3	830.5	2390.4	1670.2
MT-ND2	19073.08	0.89	0.30	2.94	3.27E-03	2.41E-02	MT	1280.9	883.1	665.6	1483.0	3684.5	2457.3
MT-ND3	7404.91	0.90	0.21	4.20	2.69E-05	4.92E-04	MT	1485.2	835.5	757.3	1852.5	2729.1	1956.1
MT-ND4	43190.14	0.59	0.30	1.98	4.75E-02	1.68E-01	MT	2079.0	1966.3	1495.2	1606.2	4493.2	4292.1
MT-ND4L	11848.61	0.51	0.30	1.71	8.65E-02	2.52E-01	MT	4094.0	3915.9	2974.8	3111.7	8844.9	7911.8
MT-ND5	15985.10	0.28	0.32	0.89	3.73E-01	6.20E-01	MT	623.8	623.6	395.3	295.8	1117.4	1053.0
MT-ND6	1792.93	0.77	0.32	2.44	1.49E-02	7.44E-02	MT	87.3	253.6	158.8	233.0	737.1	596.2
NDUFA1	5169.87	-0.20	0.15	-1.31	1.89E-01	4.20E-01	X	488.0	560.9	693.3	489.4	574.6	496.9
NDUFA10	3854.04	-0.20	0.11	-1.79	7.35E-02	2.27E-01	2	71.9	67.2	59.8	60.7	57.8	60.6
NDUFA11	783.56	-0.09	0.16	-0.54	5.88E-01	7.82E-01	19	159.7	186.9	155.5	152.7	159.6	134.9
NDUFA12	1683.81	-0.29	0.12	-2.48	1.32E-02	6.84E-02	12	92.4	128.3	119.6	116.9	102.0	83.5
NDUFA13	290.89	-0.14	0.20	-0.69	4.88E-01	7.14E-01	19	188.8	125.6	148.1	87.0	67.4	127.3
NDUFA2	1196.06	-0.33	0.16	-2.07	3.88E-02	1.46E-01	5	103.4	127.1	102.8	72.7	96.9	104.3
NDUFA3	669.02	-0.35	0.14	-2.49	1.29E-02	6.70E-02	19	86.3	82.9	81.1	80.1	59.8	64.1
NDUFA5	3622.31	-0.16	0.13	-1.23	2.18E-01	4.59E-01	7	171.6	137.0	136.7	186.6	143.0	128.6
NDUFA6	1853.11	-0.39	0.13	-3.00	2.73E-03	2.10E-02	22	154.8	130.7	94.8	89.0	79.1	69.1
NDUFA7	62.51	-0.11	0.28	-0.38	7.08E-01	8.55E-01	19	116.6	88.6	107.9	95.1	99.9	105.2
NDUFA8	1808.99	-0.29	0.12	-2.43	1.52E-02	7.54E-02	9	179.1	164.8	172.5	167.5	136.7	136.2
NDUFA9	2370.33	-0.36	0.12	-3.12	1.80E-03	1.52E-02	12	39.2	45.5	39.2	37.6	35.6	37.7
NDUFAB1	3724.12	-0.57	0.15	-3.80	1.47E-04	2.01E-03	16	522.7	500.7	387.9	315.1	292.8	294.6
NDUFB1	686.69	-0.23	0.18	-1.28	2.02E-01	4.37E-01	14	186.1	43.8	35.5	99.7	37.5	39.5
NDUFB10	1681.63	-0.60	0.15	-4.01	6.19E-05	9.85E-04	16	132.4	121.0	104.4	94.8	78.6	76.1
NDUFB11	2986.01	-0.35	0.10	-3.38	7.19E-04	7.27E-03	X	230.4	214.8	217.3	187.1	175.9	175.0
NDUFB2	866.43	-0.13	0.13	-0.99	3.24E-01	5.74E-01	7	98.6	70.9	85.4	93.4	92.0	69.5
NDUFB3	2635.06	-0.44	0.13	-3.45	5.66E-04	5.99E-03	2	252.8	292.9	308.4	201.1	218.8	223.1
NDUFB4	3737.93	-0.43	0.11	-3.79	1.53E-04	2.07E-03	3	220.3	197.9	204.1	174.7	181.9	145.4
NDUFB5	3886.16	-0.25	0.11	-2.21	2.71E-02	1.14E-01	3	122.3	164.4	129.1	117.9	117.4	115.6
NDUFB6	2371.89	-0.19	0.16	-1.18	2.39E-01	4.85E-01	9	141.1	181.7	152.1	132.1	161.1	136.3
NDUFB7	1092.43	-0.46	0.13	-3.61	3.10E-04	3.70E-03	19	186.4	176.6	181.0	153.8	130.9	121.2
NDUFB8	695.14	-0.18	0.14	-1.28	1.99E-01	4.34E-01	10	203.9	164.2	234.1	179.7	151.5	130.6
NDUFB9	4294.76	-0.30	0.15	-1.95	5.17E-02	1.78E-01	8	308.1	405.2	393.5	277.2	300.7	280.4

NDUFC1	2056.93	-0.30	0.12	-2.42	1.57E-02	7.70E-02	4	222.9	206.6	133.2	172.0	184.8	135.0
NDUFC2	1158.35	0.22	0.13	1.75	8.06E-02	2.41E-01	11	82.7	68.8	71.7	84.4	84.7	85.5
NDUFS1	7314.38	-0.04	0.14	-0.28	7.82E-01	8.96E-01	2	68.7	48.7	55.4	54.1	43.5	62.8
NDUFS2	3763.70	-0.84	0.14	-6.02	1.77E-09	1.05E-07	1	233.8	180.1	220.4	130.2	102.7	127.9
NDUFS3	1989.39	-0.23	0.13	-1.77	7.74E-02	2.35E-01	11	110.2	110.8	90.4	104.6	92.0	87.9
NDUFS4	1350.04	0.13	0.18	0.71	4.79E-01	7.07E-01	5	100.0	145.7	118.3	129.4	145.2	147.2
NDUFS5	5085.17	-0.31	0.10	-3.09	2.00E-03	1.65E-02	1	782.7	734.2	763.7	674.4	637.2	594.6
NDUFS6	1471.07	-0.33	0.18	-1.87	6.15E-02	2.01E-01	5	137.4	85.8	100.9	81.0	79.4	54.9
NDUFS7	1271.26	-0.44	0.13	-3.29	1.01E-03	9.56E-03	19	55.4	52.2	61.8	42.8	42.2	35.1
NDUFS8	1542.02	-0.11	0.12	-0.99	3.23E-01	5.74E-01	11	136.8	93.2	100.1	117.6	115.0	107.5
NDUFV1	5627.86	-0.36	0.13	-2.83	4.59E-03	3.10E-02	11	197.6	264.8	152.7	178.1	139.4	199.7
NDUFV2	872.22	-0.04	0.15	-0.29	7.72E-01	8.91E-01	18	293.7	288.2	320.7	306.9	314.0	305.2
NDUFV3	1224.26	-0.21	0.13	-1.62	1.05E-01	2.88E-01	21	37.8	33.5	33.9	38.3	32.5	31.4

Complex II subunits

								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
SDHA	4506.50	-0.12	0.13	-0.95	3.43E-01	5.92E-01	5	150.3	124.8	144.7	151.0	113.2	135.2
SDHB	4393.03	-0.25	0.11	-2.17	2.98E-02	1.22E-01	1	298.9	306.8	273.6	284.5	254.7	239.7
SDHC	8704.97	-0.66	0.12	-5.31	1.11E-07	4.07E-06	1	158.0	91.1	108.2	53.6	56.6	60.3
SDHD	3186.06	-0.12	0.12	-0.99	3.24E-01	5.74E-01	11	269.8	246.9	248.5	272.0	221.7	234.2

Complex III subunits

								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
UQCRQ	2546.42	-0.16	0.13	-1.21	2.24E-01	4.67E-01	5	235.1	266.5	216.1	244.5	211.8	212.6
UQCRH	10329.24	-0.40	0.13	-3.13	1.74E-03	1.48E-02	1	1263.1	1330.1	1313.7	941.7	1133.5	897.4
UQCRFS1	3513.30	-0.17	0.12	-1.44	1.50E-01	3.61E-01	19	85.1	70.4	86.6	83.4	67.6	71.4
UQCRC2	7956.88	-0.12	0.12	-0.99	3.24E-01	5.75E-01	16	231.8	244.5	221.1	238.8	311.6	273.8
UQCRC1	4811.41	-0.39	0.11	-3.53	4.20E-04	4.75E-03	3	204.8	175.9	186.4	174.1	143.9	153.0
UQCRB	9783.29	-0.13	0.10	-1.33	1.85E-01	4.14E-01	8	269.1	360.1	277.9	283.4	353.8	344.4
UQCR11	1516.21	-0.47	0.12	-3.93	8.52E-05	1.26E-03	19	148.3	154.1	162.9	115.3	110.1	123.2
UQCR10	2950.82	-0.42	0.12	-3.49	4.84E-04	5.33E-03	22	348.0	313.5	297.1	269.8	249.5	227.5

Complex IV subunits													
								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adj p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
MT-CYB	19085.01	0.21	0.23	0.93	3.55E-01	6.03E-01	MT	1301.1	1161.1	894.8	804.3	1811.4	1472.4
MT-CO3	48790.44	0.49	0.22	2.27	2.33E-02	1.03E-01	MT	3951.6	3754.4	3426.2	3284.8	6954.6	6707.8
MT-CO2	45726.96	0.50	0.22	2.26	2.41E-02	1.05E-01	MT	4758.7	3713.7	3374.9	3543.8	7463.8	7323.0
MT-CO1	74446.81	0.52	0.30	1.74	8.18E-02	2.43E-01	MT	2804.1	3153.3	2582.6	2183.3	5867.1	7079.4
COX8A	3222.50	-0.23	0.14	-1.68	9.30E-02	2.65E-01	11	671.2	551.2	586.0	607.1	523.6	442.0
COX7C	6852.05	-0.20	0.10	-2.00	4.50E-02	1.62E-01	5	419.1	455.1	432.4	424.1	394.4	390.5
COX7B	10523.00	-0.23	0.16	-1.42	1.55E-01	3.68E-01	X	229.5	266.4	313.7	260.7	279.9	252.3
COX7A2L	3746.83	0.05	0.15	0.33	7.40E-01	8.74E-01	2	116.5	97.2	197.9	97.0	186.4	231.5
COX7A2	4122.28	-0.29	0.16	-1.79	7.38E-02	2.27E-01	6	468.3	603.7	506.9	360.4	465.7	492.5
COX6C	5872.15	-0.28	0.18	-1.54	1.23E-01	3.18E-01	8	421.8	691.4	559.5	426.2	574.9	509.1
COX6B1	4914.69	-0.38	0.11	-3.48	5.01E-04	5.46E-03	19	595.2	544.9	536.5	497.6	416.9	434.1
COX6A1	5094.93	-0.33	0.11	-3.04	2.35E-03	1.86E-02	12	725.1	670.8	741.5	648.4	538.4	629.1
COX5B	2862.01	-0.46	0.15	-3.12	1.81E-03	1.52E-02	2	368.8	320.7	287.5	257.6	265.9	194.0
COX5A	6853.96	-0.62	0.10	-6.09	1.11E-09	6.93E-08	15	506.6	468.3	458.4	327.3	309.7	300.2
COX4I1	7072.12	-0.32	0.10	-3.10	1.95E-03	1.62E-02	16	480.4	348.4	363.0	301.6	294.1	284.5

Complex V subunits													
								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
ATP5A1	30437.53	-0.26	0.12	-2.14	3.25E-02	1.29E-01	18	1510.8	1099.1	1480.2	386.1	922.5	884.5
ATP5B	39671.78	-0.28	0.11	-2.54	1.10E-02	6.01E-02	12	1541.3	1413.5	1521.9	1362.6	1139.6	1222.2
ATP5C1	8610.75	-0.20	0.10	-1.97	4.87E-02	1.71E-01	10	541.5	533.0	533.4	518.1	493.4	452.2
ATP5D	658.93	-0.42	0.21	-2.02	4.37E-02	1.59E-01	19	80.5	33.6	28.9	31.4	31.4	39.8
ATP5E	7641.64	-0.19	0.12	-1.60	1.11E-01	2.97E-01	20	166.2	178.1	162.5	147.7	155.7	157.8
ATP5F1	10689.84	-0.26	0.11	-2.34	1.91E-02	8.95E-02	1	536.6	488.8	420.1	411.8	219.0	430.7
ATP5G1	2445.84	-0.55	0.11	-5.00	5.59E-07	1.74E-05	17	267.9	256.7	276.9	206.3	177.7	173.6
ATP5G2	6291.98	-0.13	0.16	-0.85	3.95E-01	6.39E-01	12	424.6	304.8	446.3	336.6	350.6	404.9
ATP5G3	8124.60	-0.42	0.14	-2.91	3.66E-03	2.65E-02	2	164.3	148.8	147.5	126.7	140.3	102.7
ATP5H	3018.03	-0.28	0.10	-2.68	7.31E-03	4.39E-02	17	432.4	397.7	427.4	384.2	350.2	343.4
ATP5I	3774.04	-0.56	0.13	-4.21	2.57E-05	4.73E-04	4	944.9	1012.9	808.5	732.5	606.3	618.0
ATP5J	4013.39	-0.41	0.11	-3.79	1.53E-04	2.07E-03	21	529.3	507.1	535.0	426.7	400.8	358.8
ATP5J2	5677.49	-0.53	0.11	-4.73	2.26E-06	5.92E-05	7	1268.6	1286.4	1178.8	894.7	893.2	809.0
ATP5L	5960.49	-0.22	0.11	-1.98	4.72E-02	1.68E-01	11	238.1	242.8	228.4	209.6	212.2	218.4
ATP5O	3654.36	-0.16	0.15	-1.08	2.80E-01	5.32E-01	21	383.5	460.3	405.4	316.5	353.9	373.6
ATP5S	756.77	0.48	0.15	3.26	1.10E-03	1.02E-02	14	7.8	8.5	8.7	10.8	10.8	11.5
ATP5SL	1434.19	-0.30	0.13	-2.38	1.75E-02	8.35E-02	19	52.3	52.2	43.8	58.1	40.4	43.2
MT-ATP6	23085.29	0.82	0.23	3.52	4.24E-04	4.78E-03	MT	685.3	541.8	420.9	622.6	1323.6	1442.8
MT-ATP8	9435.67	0.73	0.23	3.18	1.48E-03	1.30E-02	MT	6605.7	5544.1	4509.0	6287.4	12768.1	13969.6

Mitoflavoproteins													
								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
ACAD8	856.08	0.12	0.18	0.66	5.07E-01	7.28E-01	11	25.4	20.1	22.4	20.9	29.2	32.4
ACAD9	1624.74	-0.02	0.14	-0.14	8.89E-01	9.47E-01	3	55.9	49.6	43.6	51.9	51.6	55.6
ACADM	3444.27	0.11	0.14	0.76	4.46E-01	6.82E-01	1	123.3	82.5	95.5	105.8	112.1	95.5
ACADS	384.18	-0.41	0.19	-2.14	3.24E-02	1.29E-01	12	20.3	13.5	19.6	13.3	12.9	14.5
ACADSB	2908.66	0.22	0.18	1.26	2.07E-01	4.43E-01	10	28.2	39.3	37.0	60.0	32.5	38.1
ACADVL	2287.98	0.29	0.19	1.58	1.14E-01	3.03E-01	17	68.0	106.5	70.2	83.2	110.0	121.4
AIFM1	5063.42	-0.05	0.14	-0.33	7.39E-01	8.73E-01	X	160.5	134.1	194.4	171.4	161.6	160.9
COQ6	470.35	0.17	0.16	1.08	2.81E-01	5.33E-01	14	14.8	18.4	18.0	21.6	21.2	21.8
CYB5R3	1704.29	-0.14	0.12	-1.18	2.39E-01	4.85E-01	22	36.2	33.5	39.7	36.2	32.1	36.5
DHODH	526.68	-0.01	0.18	-0.05	9.61E-01	9.81E-01	16	15.4	9.6	12.6	15.4	9.2	10.9
DLD	5725.93	0.31	0.12	2.52	1.16E-02	6.25E-02	7	124.8	101.4	119.1	160.3	146.1	135.2
DMGDH	40.39	0.89	0.31	2.86	4.29E-03	2.97E-02	5	0.6	1.7	1.7	3.9	2.8	3.6
ETFA	9585.29	-0.14	0.11	-1.25	2.10E-01	4.48E-01	15	348.3	297.6	289.6	337.4	312.5	312.0
ETFDH	868.33	-0.24	0.18	-1.29	1.98E-01	4.33E-01	4	27.9	16.9	26.8	23.8	22.0	17.6
FDXR	352.22	-0.08	0.17	-0.45	6.50E-01	8.22E-01	17	14.3	13.5	13.6	13.9	13.4	13.2
FOXRED1	1623.77	-0.05	0.12	-0.40	6.89E-01	8.45E-01	11	50.9	53.9	50.5	59.2	65.7	56.5
GCDH	511.69	0.23	0.18	1.31	1.91E-01	4.22E-01	19	19.6	19.8	16.0	25.0	22.1	25.4
GPD2	2921.79	0.11	0.14	0.79	4.30E-01	6.68E-01	2	35.5	29.0	41.3	48.6	42.3	35.9
GSR	3226.33	0.13	0.17	0.73	4.62E-01	6.95E-01	8	71.5	56.2	86.7	97.7	82.5	63.1
IVD	2008.41	0.22	0.14	1.55	1.22E-01	3.17E-01	15	53.0	61.8	41.6	70.0	72.8	65.4
L2HGDH	1740.12	-0.31	0.12	-2.50	1.25E-02	6.58E-02	14	29.5	23.0	26.9	25.0	21.6	20.0
MAOA	1828.29	0.13	0.12	1.08	2.80E-01	5.32E-01	X	30.1	24.1	34.4	35.8	31.4	31.1
NDUFV1	5627.86	-0.36	0.13	-2.83	4.59E-03	3.10E-02	11	197.6	264.8	152.7	178.1	139.4	199.7
SDHA	4506.50	-0.12	0.13	-0.95	3.43E-01	5.92E-01	5	150.3	124.8	144.7	151.0	113.2	135.2
PPOX	608.61	-0.34	0.22	-1.57	1.16E-01	3.06E-01	1	30.8	41.9	35.4	18.3	30.6	36.5
PRODH	7.96	-0.03	0.28	-0.10	9.22E-01	9.64E-01	22	20.3	23.0	18.2	14.9	15.9	21.3
SQRDL	56.60	0.61	0.30	2.07	3.80E-02	1.45E-01	15	2.0	2.8	3.3	4.3	4.9	3.9
TXNRD2	739.38	-0.17	0.14	-1.17	2.41E-01	4.87E-01	22	26.0	28.3	29.2	28.9	26.1	25.8

Flavin-handling proteins														
									Transcripts per million (TPM)					
Gene	**base Mean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**Mito. localisation (IMPI)**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
FLAD1	2173.18	-1.06	0.15	-7.06	1.65E-12	1.78E-10	1	*Known*	93.2	83.0	64.7	61.0	36.2	43.0
RFK	1486.17	0.09	0.15	0.59	5.56E-01	7.61E-01	9	*Non-mito.*	41.7	40.1	35.6	48.4	47.5	35.1
SLC52A2	676.41	-0.12	0.18	-0.68	4.99E-01	7.23E-01	8	*Non-mito.*	17.8	28.4	19.2	26.8	16.7	19.9

Folate-metabolising proteins														
									Transcripts per million (TPM)					
Gene	**base Mean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**Mito. localisation (IMPI)**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
MTHFD1	9389.93	-0.02	0.13	-0.16	8.73E-01	9.40E-01	14	*Non-mito.*	181.3	162.6	189.0	206.5	205.5	159.0
MTHFD1L	3596.58	0.01	0.15	0.08	9.34E-01	9.69E-01	6	*Known*	108.5	68.4	89.5	135.3	97.3	87.8
MTHFD2	19033.36	0.23	0.17	1.35	1.78E-01	4.04E-01	2	*Known*	177.3	198.7	297.3	332.1	254.8	275.0
MTHFD2L	708.60	0.45	0.15	3.10	1.95E-03	1.61E-02	4	*Known*	16.4	12.1	16.3	20.0	20.1	29.0
MTHFR	449.23	-0.10	0.16	-0.64	5.22E-01	7.40E-01	1	*Non-mito.*	5.8	5.6	6.1	4.7	4.9	6.3
MTHFS	301.85	0.14	0.21	0.66	5.06E-01	7.28E-01	15	*Known*	51.5	62.9	43.3	94.6	53.0	39.9
DHFR	8864.72	-0.12	0.13	-0.90	3.68E-01	6.16E-01	5	*Non-mito.*	260.1	149.3	184.7	200.7	138.8	214.5
SHMT1	1697.15	0.28	0.13	2.19	2.83E-02	1.17E-01	17	*Non-mito.*	63.4	54.4	57.6	85.9	67.8	78.1
SHMT2	11316.49	0.39	0.16	2.53	1.15E-02	6.20E-02	12	*Known*	261.3	342.6	377.9	461.2	351.6	463.4

50 most differentially expressed genes between HAP1 and SLCKO cells													
								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
ALDH1A2	2242.49	-5.26	0.22	-24.26	4.63E-130	5.54E-126	15	81.4	91.7	79.6	1.3	2.0	0.9
FOLH1	650.04	-4.75	0.22	-21.30	1.16E-100	6.96E-97	11	38.0	39.3	33.0	2.2	0.7	0.8
KLF4	1135.96	-3.70	0.21	-17.37	1.31E-67	5.24E-64	9	84.6	44.6	33.6	3.2	4.3	3.1
DPP4	432.42	-2.83	0.18	-15.67	2.25E-55	6.72E-52	2	16.4	19.2	14.3	3.5	2.7	2.3
ZNF544	520.01	-3.45	0.24	-14.43	3.14E-47	7.53E-44	19	10.4	18.9	10.9	1.0	2.8	2.4
TUBB4A	4105.52	-2.13	0.15	-14.28	2.90E-46	5.78E-43	19	171.9	199.1	146.6	41.0	40.6	39.7
PSMB8	907.56	-2.27	0.16	-13.90	6.71E-44	1.15E-40	6	84.1	101.2	86.5	16.5	19.4	16.2
TPM2	1093.53	3.23	0.24	13.71	9.44E-43	1.41E-39	9	6.9	4.4	6.0	73.3	126.4	59.2
ZNF788	763.83	-3.30	0.24	-13.51	1.34E-41	1.79E-38	19	27.1	34.3	23.4	2.0	2.2	0.9
CD70	1160.09	-2.35	0.17	-13.46	2.60E-41	3.11E-38	19	195.7	221.9	170.9	34.5	30.0	42.0
HRASLS5	480.48	3.29	0.25	13.38	7.86E-41	8.55E-38	11	1.4	0.7	2.5	20.4	23.1	27.7
GTF2H2	200.79	-3.28	0.25	-13.37	9.43E-41	9.40E-38	5	33.1	39.7	32.4	4.6	5.7	3.7
MEST	526.64	3.46	0.26	13.06	5.89E-39	5.42E-36	7	1.8	1.2	1.7	22.2	52.9	22.4
ZNF329	496.44	-2.90	0.23	-12.72	4.47E-37	3.82E-34	19	8.1	9.5	5.7	1.3	2.0	2.4
HKR1	1360.92	-1.88	0.15	-12.28	1.19E-34	9.52E-32	19	59.8	32.9	45.5	16.9	14.2	14.5
FAM213A	677.80	-2.23	0.19	-11.68	1.58E-31	1.18E-28	10	33.5	32.3	32.5	4.8	7.6	4.9
NLRC5	718.24	-1.80	0.16	-11.63	2.93E-31	2.06E-28	16	22.6	28.8	17.4	6.8	7.1	7.8
TAGLN2	10615.83	-1.24	0.11	-11.61	3.58E-31	2.38E-28	1	624.9	679.0	631.8	296.6	263.9	282.0
CLDN11	440.28	3.05	0.27	11.40	4.09E-30	2.58E-27	3	3.2	1.3	0.9	24.6	30.1	33.2
PRLR	2528.41	-2.13	0.19	-11.25	2.45E-29	1.47E-26	5	21.5	31.0	36.1	6.0	7.5	4.5
ZNF611	897.99	-1.66	0.16	-10.62	2.41E-26	1.37E-23	19	19.1	27.6	24.7	8.1	8.1	6.9
MYBPC1	500.00	-2.58	0.25	-10.47	1.22E-25	6.65E-23	12	28.5	47.7	28.9	4.9	5.1	1.7
ADGRG6	704.80	2.04	0.20	10.42	1.97E-25	1.03E-22	6	3.4	2.9	3.5	11.6	17.8	11.7
PCDHB3	291.90	-2.26	0.22	-10.25	1.19E-24	5.94E-22	5	13.3	12.8	9.2	2.2	1.5	2.3
KIAA1551	7765.63	-1.70	0.17	-10.10	5.50E-24	2.63E-21	12	113.5	132.9	92.0	32.9	39.3	35.6
CD320	256.97	2.67	0.27	9.78	1.40E-22	6.44E-20	19	2.0	2.2	2.0	28.6	38.4	10.3
SCGB3A2	341.13	-2.63	0.27	-9.63	6.08E-22	2.69E-19	5	104.8	81.7	82.7	4.0	13.7	6.6
C3	145.10	-2.70	0.28	-9.47	2.84E-21	1.21E-18	19	5.9	6.8	6.8	0.5	1.7	1.0
A4GALT	712.71	-1.59	0.17	-9.46	3.00E-21	1.24E-18	22	39.9	39.7	31.1	12.1	10.8	14.9
CHST2	936.71	-2.19	0.23	-9.46	3.18E-21	1.27E-18	3	33.8	27.4	15.1	5.3	3.4	4.4

								Transcripts per million (TPM)					
Gene	**baseMean**	**Log_2 (FoldChange)**	**lfcSE**	**stat**	**p-value**	**Adjusted p-value**	**Chr.**	**HAP1 A**	**HAP1 B**	**HAP1 C**	**SLCKO A**	**SLCKO B**	**SLCKO C**
GREM2	302.12	-2.37	0.25	-9.37	7.49E-21	2.89E-18	1	9.9	12.5	5.8	1.1	1.1	1.5
ZNF701	192.81	-2.32	0.25	-9.35	8.43E-21	3.15E-18	19	8.2	12.1	7.3	1.0	1.1	1.1
FLVCR2	637.48	-1.77	0.19	-9.32	1.14E-20	4.10E-18	14	26.6	21.6	17.1	5.5	7.1	5.5
PIK3AP1	2778.00	-1.56	0.17	-9.32	1.17E-20	4.10E-18	10	77.7	65.2	53.3	25.3	16.7	22.4
TCEAL9	2139.46	2.82	0.30	9.32	1.20E-20	4.10E-18	X	14.2	5.4	18.0	266.3	316.8	322.4
GLIPR2	781.01	-1.88	0.20	-9.29	1.56E-20	5.18E-18	9	37.6	64.6	46.1	11.4	12.3	11.5
KCNK1	880.05	-2.17	0.24	-9.16	5.13E-20	1.66E-17	1	44.7	62.4	47.1	13.8	5.3	7.5
ASPHD1	285.08	1.80	0.20	9.08	1.06E-19	3.34E-17	16	5.1	5.1	6.9	24.4	24.3	18.6
FEZ2	2963.56	-1.03	0.11	-9.02	1.92E-19	5.90E-17	2	69.1	66.2	63.9	41.3	32.0	37.9
MOCOS	619.98	1.68	0.19	8.88	6.68E-19	2.00E-16	18	2.9	3.0	4.5	11.6	15.3	9.7
PPP1R3B	807.50	-2.00	0.23	-8.84	9.47E-19	2.76E-16	8	20.2	19.3	12.9	3.0	2.8	4.9
MLF1	519.35	-2.42	0.27	-8.83	1.05E-18	2.98E-16	3	17.8	31.8	21.0	2.0	5.0	3.0
COL5A2	1110.47	2.70	0.31	8.72	2.70E-18	7.51E-16	2	0.8	0.6	1.6	24.3	17.9	15.6
HLA-B	5700.95	-1.19	0.14	-8.57	1.02E-17	2.77E-15	6	395.5	450.9	407.6	134.9	202.2	186.6
TTYH1	871.01	-2.24	0.26	-8.53	1.48E-17	3.95E-15	19	55.5	65.3	66.6	4.1	16.6	8.8
CA8	191.64	-2.18	0.26	-8.51	1.67E-17	4.35E-15	8	8.3	10.2	5.4	1.4	1.5	1.5
TNFSF9	353.33	-1.93	0.23	-8.41	4.27E-17	1.09E-14	19	20.4	34.5	20.3	5.5	5.6	5.1
PLA2G4A	1999.12	-1.15	0.14	-8.33	8.28E-17	2.06E-14	1	68.6	75.5	73.0	31.5	36.4	29.8
KLF3	2116.76	-1.00	0.12	-8.31	9.45E-17	2.31E-14	4	41.4	42.5	38.4	21.5	20.7	20.6
FOXE1	207.35	-1.97	0.24	-8.26	1.46E-16	3.50E-14	9	6.2	9.7	7.2	1.5	1.3	1.8

Appendix 7.8 Complexome data for conditioned SLC25A32 KO cells and conditioned HAP1 cells

7.8A Table of proteins used in mass estimation of HAP1 mitochondrial sample.

Protein	Mass (kDa)	Native state	Total Mass (kDa)	Gel slice
Mitochondrial ribosome (39S + 28S)	-	39S + 28S subunits	2900	11
Complex V (ATP synthase)	770	monomer	770	25
Complex III (Cytochrome bc-1 complex)	245	dimer	490	30
Trifunctional enzyme	α - 80, β - 51	hetero-octamer	485	31
Isocitrate (NAD) dehydrogenase	α - 40, β - 42, γ - 43 (2:1:1, heterotetramer = 165)	hetero-octamer	330	32
NADP transhydrogenase	114	dimer	228	36
Complex IV (cytochrome c oxidase)	200	monomer	200	40
Complex II (succinate dehydrogenase)	130	monomer	130	48
Mitochondrial intermediate peptidase	80	monomer	80	50
Arginase-2	38	monomer	114	51
SLC25A13 (citrin)	74	monomer	74	53
SLC25A12 (aralar 1)	75	monomer	75	53
Electron transfer protein	α - 35, β - 28	heterodimer	63	54
Citrate synthase	52	dimer	104	55
Acylpyruvase	24	dimer	48	55
NADH-cytochrome b5 reductase 3	34	monomer	34	59
Proline dehydrogenase	68	monomer	68	59
Electron transfer ubiquinone oxidoreductase	69	monomer	69	60
NADPH:adrenodoxin oxidoreductase	54	monomer	54	62
2-oxoglutarate/malate carrier protein	34	dimer	68	62

7.8B Table of proteins used in mass estimation of SLC25A32 knockout mitochondrial sample.

Protein	Mass (kDa)	Native state	Total Mass (kDa)	Gel slice
Mitochondrial ribosome (39S + 28S)	-	39S + 28S subunits	2900	10
Propionyl-CoA carboxylase	α -80 and β - 58	heterododecamer	820	20
Complex V (ATP synthase)	770	monomer	770	23
Complex III (Cytochrome bc-1 complex)	245	dimer	490	28
Isocitrate (NAD) dehydrogenase	α - 40, β - 42, γ - 43 (2:1:1, heterotetramer = 165)	hetero-octamer	330	31
NADP transhydrogenase	114	dimer	228	34
Complex IV (cytochrome c oxidase)	200	monomer	200	39
SLC25A13 (citrin)	74	monomer	74	40
SLC25A12 (aralar 1)	75	monomer	75	40
Arginase-2	38	monomer	114	49
Mitochondrial intermediate peptidase	80	monomer	80	50

Citrate synthase	52	dimer	104	51
Dimethyladenosine transferase 2	45	monomer	45	52
Sulphite oxidase	60	dimer	120	58
Folylpolyglutamate synthase	64	monomer	64	59
ADP/ATP translocase (adenine nucleotide translocase)	30	monomer	30	59
Electron transfer ubiquinone oxidoreductase	69	monomer	69	59
NADH-cytochrome b5 reductase 3	34	monomer	34	59
NADPH:adrenodoxin oxidoreductase	54	monomer	54	59
Acylpyruvase	24	dimer	48	62
2-oxoglutarate/malate carrier protein	34	dimer	68	63
Adenylate kinase 2	26	monomer	26	63

7.8C Complexome data for mitochondria from conditioned HAP1 cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-ND1	11.95	3	35.6	92.81
MT-ND2	7.49	1	38.9	84.82
MT-ND4	10.02	4	51.5	45.56
MT-ND5	7.96	4	67.0	242.82
MT-ND6	8.05	1	18.6	119.95
NDUFA1	18.57	1	8.1	39.34
NDUFA10	36.62	12	40.7	345.74
NDUFA11	31.91	3	14.8	269.83
NDUFA12	61.38	8	17.1	438.88
NDUFA13	67.36	10	16.7	890.89
NDUFA2	50.51	6	10.9	288.83
NDUFA3	50.00	3	9.3	118.39
NDUFA5	43.97	5	13.5	414.57
NDUFA6	53.25	6	17.9	406.45
NDUFA7	55.75	8	12.5	723.27
NDUFA8	21.51	4	20.1	230.93
NDUFA9	67.37	16	42.5	757.27
NDUFAB1	17.95	2	17.4	158.57
NDUFB1	34.48	3	7.0	172.09
NDUFB10	62.21	11	20.8	405.27
NDUFB11	42.48	4	17.3	153.42
NDUFB3	41.84	4	11.4	220.67
NDUFB4	44.19	4	15.2	419.82
NDUFB5	31.22	5	21.7	391.27
NDUFB6	36.72	4	15.5	205.12

NDUFB7	24.09	2	16.4	541.78
NDUFB8	38.17	4	21.8	339.74
NDUFB9	55.31	5	21.8	409.76
NDUFC1	27.63	1	8.7	30.52
NDUFC2	37.82	6	14.2	106.69
NDUFS2	45.14	13	52.5	716.81
NDUFS3	48.11	12	30.2	824.76
NDUFS4	39.43	5	20.1	395.91
NDUFS5	25.47	4	12.5	189.94
NDUFS6	29.03	3	13.7	133.76
NDUFS7	23.00	5	23.5	631.25
NDUFS8	35.71	6	23.7	413.99
NDUFV1	36.85	12	50.8	613.79
NDUFV2	49.00	12	27.4	830.41
NDUFV3	22.22	1	11.9	47.19

Complex II subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
SDHA	44.73	19	72.6	252.26
SDHB	52.50	13	31.6	17.99
SDHC	8.88	1	18.6	49.43
SDHD	21.38	2	17.0	45.12

Complex III subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-CYB	2.37	1	42.7	62.90
UQCR10	38.10	2	7.3	377.71
UQCRB	58.56	7	13.5	897.52
UQCRC1	40.42	15	52.6	2042.28
UQCRC2	46.58	13	48.4	2184.56
UQCRFS1	45.99	11	29.6	1097.92
UQCRH	12.09	1	10.7	23.96
UQCRQ	36.59	4	9.9	151.76

Complex IV subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
COX4I1	55.03	9	19.6	1274.03
COX5A	38.67	6	16.8	796.31
COX5B	46.51	6	13.7	615.10

COX6A1	16.51	1	12.1	69.35
COX6B1	31.40	3	10.2	116.25
COX6C	52.00	7	8.8	559.30
COX7A2	37.35	2	9.4	294.16
COX7C	25.40	1	7.2	37.80
COX8A	27.54	1	7.6	111.91
MT-CO1	7.60	3	57.0	169.95
MT-CO2	27.75	5	25.5	736.43
MT-CO3	3.45	1	29.9	31.97

Complex V subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
ATP5A1	66.18	36	59.7	17329.07
ATP5B	77.69	30	56.5	15579.59
ATP5F1	54.69	15	28.9	3126.85
ATP5O	59.62	12	23.3	2427.68
ATP5H	72.67	10	18.5	1830.72
ATP5C1	34.90	12	33.0	1402.16
ATP5L	47.57	5	11.4	1289.88
ATP5J	53.70	7	12.6	989.00
ATP5I	66.67	6	7.9	954.78
ATP5J2	39.36	3	10.9	805.97
ATP5D	23.81	3	17.5	548.77
ATP5E	31.37	2	5.8	181.37
MT-ATP6	6.19	2	24.8	166.48
MT-ATP8	36.76	2	8.0	106.41

7.8D Complexome data for mitochondria from conditioned SLC25A32 knockout cells. Score threshold for 95% confidence interval is 21.

Complex I subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-ND1	11.95	3	35.6	111.11
MT-ND2	2.59	1	38.9	79.27
MT-ND4	8.93	3	51.5	187.37
MT-ND5	7.96	4	67.0	423.21
MT-ND6	8.05	1	18.6	124.03
NDUFA1	18.57	1	8.1	51.08
NDUFA10	31.83	13	40.7	433.47
NDUFA11	31.91	3	14.8	243.74

NDUFA12	46.90	6	17.1	312.57
NDUFA13	67.36	10	16.7	1018.67
NDUFA2	42.42	4	10.9	199.82
NDUFA3	50.00	3	9.3	80.45
NDUFA5	43.97	5	13.5	361.40
NDUFA6	44.16	7	17.9	311.90
NDUFA7	55.75	8	12.5	194.57
NDUFA8	21.51	4	20.1	248.06
NDUFA9	61.01	19	42.5	1007.59
NDUFAB1	16.67	2	17.4	91.03
NDUFB1	34.48	3	7.0	135.93
NDUFB10	62.21	12	20.8	521.55
NDUFB11	34.64	4	17.3	206.71
NDUFB3	28.57	3	11.4	307.42
NDUFB4	37.21	4	15.2	315.70
NDUFB5	31.22	7	21.7	280.74
NDUFB6	36.72	4	15.5	387.85
NDUFB7	31.39	3	16.4	618.25
NDUFB8	36.02	4	21.8	299.38
NDUFB9	29.05	4	21.8	447.14
NDUFC1	27.63	2	8.7	24.78
NDUFC2	37.82	8	14.2	147.63
NDUFS1	49.79	27	79.4	1362.51
NDUFS2	40.82	15	52.5	687.42
NDUFS3	50.38	12	30.2	1103.22
NDUFS4	31.43	5	20.1	202.27
NDUFS5	25.47	3	12.5	135.58
NDUFS6	19.35	3	13.7	46.45
NDUFS7	18.31	5	23.5	661.13
NDUFS8	28.10	5	23.7	478.53
NDUFV1	29.53	12	50.8	405.43
NDUFV2	37.35	9	27.4	308.04

Complex II subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
SDHA	35.99	18	72.6	150.42
SDHB	26.79	6	31.6	202.55
SDHC	14.20	2	18.6	66.11
SDHD	10.69	2	17.0	68.54

Complex III subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
CYC1	32.00	7	35.4	859.95
MT-CYB	2.37	1	42.7	83.21
UQCR10	38.10	2	7.3	291.46
UQCRB	47.75	7	13.5	1050.50
UQCRC1	35.00	14	52.6	1808.48
UQCRC2	35.10	11	48.4	2444.43
UQCRFS1	41.97	9	29.6	977.41
UQCRQ	18.29	2	9.9	158.75

Complex IV subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-CO1	7.60	3	57.0	173.41
MT-CO2	27.75	5	25.5	816.86
COX4I1	46.75	11	19.6	944.28
COX5A	38.67	6	16.8	755.66
COX5B	41.09	6	13.7	742.77
COX6A1	16.51	1	12.1	147.50
COX6B1	22.09	2	10.2	26.05
COX6C	44.00	7	8.8	300.47
COX7A2	27.71	2	9.4	434.35
COX7C	14.29	1	7.2	44.16
COX8A	13.04	1	7.6	118.16

Complex V subunits				
Protein	**Coverage [%]**	**Unique Peptides**	**Molecular weight [kDa]**	**Score (/21)**
MT-ATP8	36.76	2	8.0	68.71
MT-ATP6	6.19	2	24.8	195.22
ATP5A1	60.58	32	59.7	8799.92
ATP5F1	52.34	15	28.9	2423.69
ATP5B	70.32	24	56.5	10386.55
ATP5H	69.57	12	18.5	1435.37
ATP5D	18.45	2	17.5	418.45
ATP5I	62.32	5	7.9	1000.34
ATP5EP2	47.06	2	5.8	170.40
ATP5J2	39.36	3	10.9	827.35
ATP5L	47.57	5	11.4	1491.45
ATP5C1	27.85	9	33.0	1768.72
ATP5O	71.83	12	23.3	2390.34
ATP5J	46.30	4	12.6	739.63

Appendix 7.9. Variant data for human complex I subunit genes

Appendix 7.9A Subunit classification

Subunit	Encoding genome	Subunit type	Subunit	Encoding genome	Subunit type
ND1	Mitochondrial	Core	**NDUFV3**	Nuclear	Supernumerary
ND2			**NDUFS4**		
ND3			**NDUFS5**		
ND4			**NDUFS6**		
ND4L			**NDUFA1**		
ND5			**NDUFA2**		
ND6			**NDUFA3**		
NDUFV1	Nuclear		**NDUFA5**		
NDUFV2			**NDUFA6**		
NDUFS1			**NDUFA7**		
NDUFS2			**NDUFA8**		
NDUFS3			**NDUFA9**		
NDUFS7			**NDUFA10**		
NDUFS8			**NDUFA11**		
			NDUFA12		
			NDUFA13		
			NDUFAB1-α/β		
			NDUFB1		
			NDUFB2		
			NDUFB3		
			NDUFB4		
			NDUFB5		
			NDUFB6		
			NDUFB7		
			NDUFB8		
			NDUFB9		
			NDUFB10		
			NDUFB11		
			NDUFC1		
			NDUFC2		

Appendix 7.9B Complete data for human complex I subunit gene variants. Residue number refers to the mature protein (distances from key residues in Appendix 7.9C).

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
289	LEU	ND1	Fully	Non-polar	no	1	L-M	2	L-M, L-Q	TMH	Buried
45	SER	ND3	Not	Polar non-charged	no	1	S-P	2	S-A, S-Y, S-F	Loop	Interface
340	ARG	ND4	Fully	Positively charged	no	1	R-H, R-S	2	R-H	Loop	Buried
64	MET	ND6	Strongly	Non-polar	no	1	M-I, M-V	2	M-I, M-V	TMH	Interface
52	ALA	ND1	Fully	Non-polar	no	1	A-T	3	A-T	Helix	Surface
110	SER	ND1	Weakly	Polar non-charged	no	1	S-N	3	S-N	TMH	Interface
132	ALA	ND1	Strongly	Non-polar	no	1	A-T	3	A-T	TMH	Buried
34	SER	ND3	Not	Polar non-charged	no	1	S-P	3	S-A, S-Y, S-T, S-C, S-F	Loop	Interface
47	ALA	ND3	Fully	Non-polar	no	1	A-T	3	A-T	Loop	Interface
36	GLY	ND6	Fully	Non-polar	no	1	G-S	3	G-S	TMH	Buried
72	ALA	ND6	Fully	Non-polar	no	1	A-V	3	A-P, A-V	TMH	Interface
259	MET	NDUFS2	Fully	Non-polar	no	1	M-T	3	M-T	Loop	Buried
393	ALA	NDUFS2	Fully	Non-polar	no	1	A-S	3	A-S	Loop	Buried
24	GLU	ND1	Fully	Negatively charged	yes	1	E-K			TMH	Buried
131	GLY	ND1	Fully	Non-polar	no	1	G-S			TMH	Buried
143	GLU	ND1	Fully	Negatively charged	no	1	E-K, E-Q			TMH	Buried
195	ARG	ND1	Fully	Positively charged	no	1	R-Q			Loop	Buried
65	VAL	ND4L	Fully	Non-polar	no	1	V-A			TMH	Interface
124	PHE	ND5	Fully	Aromatic	no	1	F-L			TMH	Buried
393	ASP	ND5	Fully	Negatively charged	no	1	D-N, D-G			TMH	Buried
60	LEU	ND6	Fully	Non-polar	no	1	L-S			TMH	Interface
63	MET	ND6	Fully	Non-polar	no	1	M-V			TMH	Buried
8	GLY	NDUFA1	Fully	Non-polar	no	1	G-R			TMH	Interface
37	ARG	NDUFA1	Fully	Positively charged	yes	1	R-S			Loop	Interface
107	GLN	NDUFA10	Fully	Polar non-charged	no	1	Q-R			Helix	Buried
286	ARG	NDUFA9	Fully	Positively charged	yes	1	R-P			Helix	Buried
208	LEU	NDUFS1	Fully	Non-polar	no	1	L-V			Beta sheet	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
218	ARG	NDUFS1	Fully	Positively charged	no	1	R-W			Loop	Interface
229	ASP	NDUFS1	Fully	Negatively charged	no	1	D-G			Loop	Buried
385	ARG	NDUFS1	Fully	Positively charged	yes	1	R-C			Loop	Buried
572	THR	NDUFS1	Fully	Polar non-charged	no	1	T-A			Beta sheet	Buried
596	ASP	NDUFS1	Fully	Negatively charged	yes	1	D-N			Helix	Buried
195	ARG	NDUFS2	Fully	Positively charged	no	1	R-Q			Loop	Buried
196	PRO	NDUFS2	Fully	Polar non-charged	no	1	P-Q			Loop	Buried
380	SER	NDUFS2	Fully	Polar non-charged	no	1	S-P			Beta sheet	Buried
109	THR	NDUFS3	Fully	Polar non-charged	no	1	T-I			Beta sheet	Buried
163	ARG	NDUFS3	Fully	Positively charged	yes	1	R-W			Loop	Buried
87	CYS	NDUFS6	Fully	Polar non-charged	no	1	C-Y			Loop	Buried
84	VAL	NDUFS7	Fully	Non-polar	no	1	V-M			Beta sheet	Buried
29	GLU	NDUFS8	Fully	Negatively charged	yes	1	E-Q			Loop	Interface
43	ARG	NDUFS8	Fully	Positively charged	no	1	R-W			Loop	Interface
45	PRO	NDUFS8	Fully	Polar non-charged	no	1	P-L			Loop	Interface
51	PRO	NDUFS8	Fully	Polar non-charged	no	1	P-L			Loop	Interface
68	ARG	NDUFS8	Fully	Positively charged	no	1	R-H			Beta sheet	Buried
104	ARG	NDUFS8	Fully	Positively charged	no	1	R-H			Loop	Interface
125	ALA	NDUFS8	Fully	Non-polar	no	1	A-D			Helix	Interface
97	ALA	NDUFV1	Weakly	Non-polar	no	1	A-T			Loop	Buried
102	PRO	NDUFV1	Fully	Polar non-charged	no	1	P-L			Loop	Interface
186	CYS	NDUFV1	Fully	Polar non-charged	no	1	C-G			Loop	Buried
403	THR	NDUFV1	Fully	Polar non-charged	no	1	T-M			Loop	Buried
30	TYR	ND1	Strongly	Aromatic	no	2	Y-H, Y-C	1	Y-H, Y-C, Y-S	TMH	Interface
164	THR	ND1	Not	Polar non-charged	no	2	T-A	1	T-A, T-S	Helix	Buried
240	THR	ND1	Not	Polar non-charged	no	2	T-M	1	T-A, T-S, T-M	TMH	Buried
1	MET	ND1	Fully	Non-polar	no	2	M-K	2	M-T, M-L, M-K	Loop	Surface
31	MET	ND1	Not	Non-polar	no	2	M-V, M-I, M-T	2	M-T, M-V, M-I	TMH	Interface
277	TYR	ND1	Fully	Aromatic	no	2	Y-C	2	Y-C, Y-H	Loop	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
57	ILE	ND2	Not	Non-polar	no	2	I-M	2	I-M, I-V, I-T	TMH	Interface
60	ILE	ND3	Fully	Non-polar	no	2	I-T	2	I-M, I-T	TMH	Interface
109	THR	ND4	Strongly	Polar non-charged	no	2	T-A	2	T-A	TMH	Buried
165	ILE	ND4	Not	Non-polar	no	2	I-T	2	I-T, I-V	TMH	Interface
313	VAL	ND4	Not	Non-polar	no	2	V-I	2	V-I	TMH	Buried
1	MET	ND5	Not	Non-polar	no	2	M-T	2	M-T, M-V	Loop	Surface
159	TYR	ND5	Weakly	Aromatic	no	2	Y-H	2	Y-H	Loop	Interface
434	GLN	ND5	Not	Polar non-charged	no	2	Q-R	2	Q-R	Loop	Interface
132	SER	ND6	Not	Polar non-charged	no	2	S-A	2	S-A, S-L, S-W, S-P, S-T	Loop	Surface
290	ARG	NDUFS2	Strongly	Positively charged	yes	2	R-Q	2	R-Q, R-G	Loop	Surface
28	LEU	ND1	Fully	Non-polar	no	2	L-M	3	L-M	TMH	Buried
215	TYR	ND1	Fully	Aromatic	no	2	Y-H	3	Y-H, Y-N	Loop	Buried
409	TYR	ND4	Strongly	Aromatic	no	2	Y-H	3	Y-H	TMH	Buried
236	ALA	ND5	Fully	Non-polar	no	2	A-T	3	A-T, A-V	TMH	Buried
26	ILE	ND6	Strongly	Non-polar	no	2	I-M, I-T	3	I-V	TMH	Interface
117	ASN	ND6	Weakly	Polar non-charged	no	2	N-D	3	N-D, N-S	Loop	Buried
21	TRP	NDUFB3	Fully	Aromatic	no	2	L-P	3	W-R	Loop	Buried
179	ARG	NDUFV1	Fully	Positively charged	no	2	R-C	3	R-C, R-H	Beta sheet	Interface
366	ARG	NDUFV1	Not	Positively charged	no	2	R-C, R-H	3	R-C, R-H	Loop	Interface
59	GLU	ND1	Fully	Negatively charged	yes	2	E-K			Loop	Interface
208	VAL	ND1	Fully	Non-polar	no	2	V-L			Loop	Interface
214	GLU	ND1	Fully	Negatively charged	no	2	E-K			Loop	Interface
285	LEU	ND1	Fully	Non-polar	no	2	L-P			TMH	Buried
71	LEU	ND2	Strongly	Non-polar	no	2	L-M			TMH	Buried
259	GLY	ND2	Fully	Non-polar	no	2	G-S			Loop	Buried
26	GLN	ND3	Fully	Polar non-charged	no	2	Q-K			Loop	Surface
66	ASP	ND3	Fully	Negatively charged	no	2	D-N			TMH	Interface
32	CYS	ND4L	Fully	Polar non-charged	no	2	C-R			TMH	Buried
145	GLU	ND5	Fully	Negatively charged	yes	2	E-G			TMH	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
171	ALA	ND5	Fully	Non-polar	no	2	A-V			TMH	Buried
237	MET	ND5	Fully	Non-polar	no	2	M-L			Loop	Buried
250	SER	ND5	Fully	Polar non-charged	no	2	S-C			TMH	Buried
465	GLY	ND5	Fully	Non-polar	no	2	G-E			TMH	Buried
59	TYR	ND6	Fully	Aromatic	no	2	Y-C			TMH	Interface
74	ALA	ND6	Fully	Non-polar	no	2	A-V			Loop	Surface
19	PRO	NDUFA1	Fully	Polar non-charged	no	2	P-S			TMH	Interface
64	GLY	NDUFA10	Fully	Non-polar	no	2	G-E			Loop	Interface
259	LEU	NDUFA10	Strongly	Non-polar	no	2	L-P			Helix	Buried
56	ARG	NDUFA13	Fully	Positively charged	no	2	R-H			TMH	Interface
44	LYS	NDUFA2	Not	Positively charged	no	2	K-T			Helix	Surface
108	GLU	NDUFA8	Weakly	Negatively charged	no	2	E-K			Helix	Surface
92	GLU	NDUFB11	Fully	Negatively charged	yes	2	E-K			Helix	Buried
63	LEU	NDUFB9	Fully	Non-polar	no	2	Y-C			Helix	Buried
96	TYR	NDUFB9	Fully	Aromatic	no	2	V-D			Loop	Interface
48	VAL	NDUFS1	Strongly	Non-polar	no	2	G-E			Loop	Interface
143	GLY	NDUFS1	Fully	Non-polar	no	2	V-A			Loop	Surface
205	VAL	NDUFS1	Fully	Non-polar	no	2	V-A			Loop	Buried
499	GLN	NDUFS1	Fully	Polar non-charged	no	2	Y-C			Helix	Buried
672	TYR	NDUFS1	Fully	Aromatic	no	2	M-V			Loop	Buried
684	MET	NDUFS1	Fully	Non-polar	no	2	Y-C			Helix	Buried
77	ASP	NDUFS2	Fully	Negatively charged	no	2	D-V			Beta sheet	Buried
85	ARG	NDUFS2	Fully	Positively charged	no	2	R-Q			Loop	Interface
105	ARG	NDUFS2	Fully	Positively charged	no	2	R-Q			Helix	Interface
115	GLU	NDUFS2	Fully	Negatively charged	no	2	E-K			Helix	Buried
191	ALA	NDUFS2	Fully	Non-polar	no	2	A-V			Loop	Buried
300	ARG	NDUFS2	Fully	Positively charged	yes	2	R-Q			Helix	Buried
410	MET	NDUFS2	Fully	Non-polar	no	2	M-K			Loop	Interface
413	ASP	NDUFS2	Fully	Negatively charged	yes	2	D-N			Helix	Buried
187	PRO	NDUFS3	Fully	Polar non-charged	no	2	P-L			Beta sheet	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
72	TRP	NDUFS4	Fully	Aromatic	no	2	W-R			Loop	Interface
77	ASP	NDUFS4	Fully	Negatively charged	yes	2	D-H			Loop	Buried
95	PRO	NDUFS5	Not	Polar non-charged	no	2	P-S			Loop	Surface
67	ARG	NDUFS7	Fully	Positively charged	no	2	R-C			Loop	Interface
107	ARG	NDUFS7	Not	Positively charged	no	2	R-H			Loop	Buried
20	ARG	NDUFS8	Not	Positively charged	yes	2	R-W			Helix	Surface
60	ARG	NDUFS8	Fully	Positively charged	no	2	R-C			Loop	Interface
120	GLY	NDUFS8	Fully	Non-polar	no	2	G-S			Loop	Buried
36	SER	NDUFV1	Strongly	Polar non-charged	no	2	S-P			Helix	Buried
68	ARG	NDUFV1	Fully	Positively charged	no	2	R-G			Loop	Buried
91	LYS	NDUFV1	Not	Positively charged	no	2	K-E			Loop	Buried
127	ARG	NDUFV1	Strongly	Positively charged	no	2	R-W			Helix	Surface
184	TYR	NDUFV1	Fully	Aromatic	no	2	Y-C			Loop	Buried
191	ALA	NDUFV1	Fully	Non-polar	no	2	A-V			Helix	Buried
226	GLU	NDUFV1	Fully	Negatively charged	yes	2	E-K			Helix	Buried
232	PRO	NDUFV1	Fully	Polar non-charged	no	2	P-L			Helix	Buried
321	ALA	NDUFV1	Weakly	Non-polar	no	2	A-V			Loop	Buried
357	GLU	NDUFV1	Not	Negatively charged	no	2	E-K			Helix	Buried
358	SER	NDUFV1	Not	Polar non-charged	no	2	S-R			Loop	Buried
412	ALA	NDUFV1	Fully	Non-polar	no	2	A-P			Helix	Buried
177	LYS	NDUFV2	Strongly	Positively charged	yes	2	K-R			Helix	Surface
4	ALA	ND1	Not	Non-polar	no	3	A-T	1	A-T, A-V	TMH	Surface
304	TYR	ND1	Not	Aromatic	no	3	Y-H	1	Y-H	TMH	Buried
150	ASN	ND2	Strongly	Polar non-charged	no	3	N-D	1	N-D, N-S	Loop	Buried
331	ALA	ND2	Not	Non-polar	no	3	A-T, A-S	1	A-T, A-S, A-G, A-V	TMH	Interface
458	ALA	ND5	Not	Non-polar	no	3	A-T	1	A-T	TMH	Buried
544	THR	ND5	Not	Polar non-charged	no	3	T-M	1	T-A, T-M	Helix	Surface
119	ASN	ND6	Not	Polar non-charged	no	3	N-D	1	N-D, N-S	Loop	Surface
11	VAL	ND1	Strongly	Non-polar	no	3	V-A	2	V-A, V-M, V-L, V-E	TMH	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
39	VAL	ND1	Strongly	Non-polar	no	3	V-I	2	V-G, V-I	Loop	Surface
64	ALA	ND1	Not	Non-polar	no	3	A-S, A-V	2	A-S, A-T, A-V	Loop	Surface
113	VAL	ND1	Fully	Non-polar	no	3	V-A	2	V-A, V-G	TMH	Buried
144	VAL	ND1	Strongly	Non-polar	no	3	V-I	2	V-I	TMH	Interface
147	ALA	ND1	Strongly	Non-polar	no	3	A-L	2	A-L, A-T	TMH	Interface
187	ILE	ND1	Not	Non-polar	no	3	I-T	2	I-S, I-T, I-V	TMH	Buried
64	ALA	ND2	Strongly	Non-polar	no	3	A-T	2	A-S, A-T	TMH	Buried
122	THR	ND2	Not	Polar non-charged	no	3	T-A	2	T-A, T-I	Loop	Buried
10	ASN	ND3	Weakly	Polar non-charged	no	3	N-D	2	N-D	TMH	Interface
423	ILE	ND4	Not	Non-polar	no	3	I-V, I-T	2	I-V, I-T	Loop	Buried
9	THR	ND5	Not	Polar non-charged	no	3	T-A	2	T-A, T-I	TMH	Buried
21	THR	ND5	Not	Polar non-charged	no	3	T-A	2	T-A, T-I	TMH	Buried
100	ILE	ND5	Fully	Non-polar	no	3	I-V	2	I-T, I-V	TMH	Buried
267	ALA	ND5	Not	Non-polar	no	3	A-T	2	A-T	Loop	Buried
398	THR	ND5	Strongly	Polar non-charged	no	3	T-A	2	T-A, T-S	TMH	Buried
576	ILE	ND5	Not	Non-polar	no	3	I-T	2	I-T, I-V	Helix	Surface
33	ILE	ND6	Weakly	Non-polar	no	3	I-V	2	I-L, I-V, I-F	TMH	Interface
58	ILE	ND6	Strongly	Non-polar	no	3	I-T	2	I-T, I-V	TMH	Interface
171	ALA	ND6	Not	Non-polar	no	3	A-T	2	A-T, A-V	TMH	Interface
32	GLY	NDUFA1	Not	Non-polar	no	3	G-R	2	G-R	Loop	Surface
103	ARG	NDUFA11	Not	Positively charged	no	3	R-L	2	R-H, R-L	TMH	Surface
40	ARG	NDUFA9	Strongly	Positively charged	no	3	R-H, R-L	2	R-H, R-L, R-C	Helix	Interface
408	LEU	NDUFS1	Strongly	Non-polar	no	3	L-V	2	L-V	Beta sheet	Buried
2	PRO	ND1	Not	Polar non-charged	no	3	P-L	3	P-L, P-S	TMH	Surface
56	PHE	ND1	Fully	Aromatic	no	3	F-L	3	F-L, F-S	Helix	Surface
128	ALA	ND1	Strongly	Non-polar	no	3	A-T	3	A-P, A-T	TMH	Buried
176	LEU	ND1	Strongly	Non-polar	no	3	L-M	3	L-M	Loop	Surface
230	ASN	ND1	Fully	Polar non-charged	no	3	N-S	3	N-D, N-H, N-T, N-S	TMH	Buried
276	ALA	ND1	Strongly	Non-polar	no	3	A-T	3	A-T	Loop	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
328	THR	ND2	Not	Polar non-charged	no	3	T-M	3	T-A, T-P, T-M	TMH	Interface
71	ALA	ND4L	Not	Non-polar	no	3	A-T	3	A-T	TMH	Buried
99	SER	ND5	Fully	Polar non-charged	no	3	S-T	3	S-A	TMH	Buried
149	ILE	ND5	Fully	Non-polar	no	3	I-S	3	I-T, I-V	TMH	Interface
112	VAL	ND6	Not	Non-polar	no	3	V-M	3	V-M	Loop	Surface
100	LEU	NDUFA10	Fully	Non-polar	no	3	L-S	3	L-S	Helix	Buried
251	GLN	NDUFA10	Fully	Polar non-charged	no	3	Q-E	3	Q-E	Loop	Surface
50	ARG	NDUFA9	Fully	Positively charged	no	3	R-W	3	R-W	Loop	Surface
279	ILE	NDUFA9	Not	Non-polar	no	3	I-M	3	I-M	Loop	Buried
157	ARG	NDUFB9	Fully	Positively charged	yes	3	R-G	3	R-G	Loop	Surface
154	ILE	NDUFS1	Fully	Non-polar	no	3	I-M	3	I-M, I-V	Loop	Buried
181	MET	NDUFS1	Fully	Non-polar	no	3	M-I	3	M-I, M-T	Loop	Buried
483	VAL	NDUFS1	Weakly	Non-polar	no	3	V-I	3	V-G, V-I	Loop	Surface
511	VAL	NDUFS1	Fully	Non-polar	no	3	V-M	3	V-M	Loop	Buried
287	VAL	NDUFS2	Strongly	Non-polar	no	3	V-A	3	V-A	Loop	Interface
347	HIS	NDUFS2	Fully	Positively charged	no	3	H-D	3	H-D	Helix	Buried
210	ARG	NDUFS3	Fully	Positively charged	no	3	R-C	3	R-C	Loop	Interface
58	ALA	NDUFS6	Strongly	Non-polar	no	3	A-V	3	A-V	Beta sheet	Surface
60	ASP	NDUFS6	Fully	Negatively charged	no	3	D-N	3	D-N	Beta sheet	Surface
167	ARG	NDUFS7	Not	Positively charged	no	3	R-G	3	R-G, R-W, R-Q	Loop	Surface
133	ALA	NDUFV1	Not	Non-polar	no	3	A-T	3	A-T	Beta sheet	Buried
337	MET	NDUFV1	Fully	Non-polar	no	3	M-L	3	M-L, M-T	Beta sheet	Buried
12	PRO	ND1	Not	Polar non-charged	no	3	P-S			TMH	Interface
25	ARG	ND1	Fully	Positively charged	yes	3	R-Q			TMH	Buried
34	ARG	ND1	Fully	Positively charged	yes	3	R-H			Loop	Interface
38	ASN	ND1	Fully	Polar non-charged	no	3	N-D			Loop	Interface
218	GLY	ND1	Fully	Non-polar	no	3	G-D			TMH	Buried
279	ARG	ND1	Fully	Positively charged	no	3	R-Q			Loop	Interface
55	ALA	ND2	Strongly	Non-polar	no	3	A-G			TMH	Buried
60	PHE	ND2	Fully	Aromatic	no	3	F-S			TMH	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
128	LEU	ND2	Strongly	Non-polar	no	3	L-Q			TMH	Buried
158	LEU	ND4	Strongly	Non-polar	no	3	L-P			TMH	Interface
206	LYS	ND4	Fully	Positively charged	no	3	K-Q			TMH	Buried
372	THR	ND4	Not	Polar non-charged	no	3	T-N			Loop	Buried
387	SER	ND4	Strongly	Polar non-charged	no	3	S-F			TMH	Interface
412	THR	ND4	Not	Polar non-charged	no	3	T-I			TMH	Interface
25	HIS	ND4L	Fully	Positively charged	no	3	H-R			Loop	Buried
41	PHE	ND4L	Fully	Aromatic	no	3	F-C			TMH	Interface
96	VAL	ND5	Not	Non-polar	no	3	V-I			TMH	Buried
239	GLY	ND5	Fully	Non-polar	no	3	G-S			Loop	Buried
243	VAL	ND5	Fully	Non-polar	no	3	V-I			TMH	Buried
253	VAL	ND5	Fully	Non-polar	no	3	V-A			TMH	Buried
312	LEU	ND5	Fully	Non-polar	no	3	L-P			TMH	Buried
348	HIS	ND5	Fully	Positively charged	no	3	N-S			TMH	Buried
392	LYS	ND5	Fully	Positively charged	no	3	K-M			TMH	Buried
415	ALA	ND5	Strongly	Non-polar	no	3	A-G			TMH	Buried
499	LEU	ND5	Not	Non-polar	no	3	L-M			TMH	Surface
505	ASN	ND5	Not	Polar non-charged	no	3	N-H			TMH	Surface
585	LYS	ND5	Fully	Positively charged	no	3	K-N			TMH	Surface
25	PRO	ND6	Fully	Polar non-charged	no	3	P-L			TMH	Interface
79	PRO	ND6	Fully	Polar non-charged	no	3	P-S			Loop	Interface
82	TRP	ND6	Fully	Aromatic	no	3	W-R			Loop	Surface
101	GLY	ND6	Not	Non-polar	no	3	G-V			TMH	Buried
43	TYR	NDUFA1	Not	Aromatic	no	3	Y-H			Helix	Interface
173	ASP	NDUFA10	Fully	Negatively charged	no	3	D-N			Beta sheet	Surface
176	VAL	NDUFA10	Fully	Non-polar	no	3	V-I			Helix	Buried
9	TRP	NDUFA11	Not	Aromatic	no	3	W-C			Loop	Surface
77	ALA	NDUFA11	Not	Non-polar	no	3	A-T			TMH	Buried
81	GLU	NDUFA11	Not	Negatively charged	no	3	E-G			Loop	Surface
68	TYR	NDUFA2	Not	Aromatic	no	3	Y-C			Loop	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
26	ALA	NDUFA9	Fully	Non-polar	no	3	A-V			Loop	Buried
33	TYR	NDUFA9	Fully	Aromatic	no	3	Y-C			Helix	Buried
287	MET	NDUFA9	Strongly	Non-polar	no	3	M-T			Helix	Buried
303	LEU	NDUFA9	Strongly	Non-polar	no	3	L-F			Loop	Buried
60	THR	NDUFB9	Fully	Polar non-charged	no	3	T-S			Helix	Interface
39	ARG	NDUFS1	Fully	Positively charged	no	3	R-Q			Loop	Buried
230	VAL	NDUFS1	Fully	Non-polar	no	3	V-G			Loop	Buried
286	ASN	NDUFS1	Fully	Polar non-charged	no	3	N-S			Loop	Surface
432	ILE	NDUFS1	Strongly	Non-polar	no	3	I-V			Helix	Buried
452	VAL	NDUFS1	Strongly	Non-polar	no	3	V-F			Beta sheet	Buried
543	ILE	NDUFS1	Fully	Non-polar	no	3	I-F			Beta sheet	Buried
553	GLY	NDUFS1	Fully	Non-polar	no	3	G-E			Helix	Buried
577	GLU	NDUFS1	Fully	Negatively charged	no	3	E-D			Loop	Buried
680	ALA	NDUFS1	Fully	Non-polar	no	3	A-T			Helix	Buried
20	TYR	NDUFS2	Fully	Aromatic	no	3	Y-C			Loop	Interface
376	VAL	NDUFS2	Fully	Non-polar	no	3	V-M			Beta sheet	Buried
42	ARG	NDUFS4	Fully	Positively charged	no	3	R-H			Loop	Surface
49	VAL	NDUFS6	Fully	Non-polar	no	3	V-M			Loop	Surface
55	ARG	NDUFS6	Fully	Positively charged	yes	3	R-W			Loop	Surface
74	ASN	NDUFS6	Fully	Polar non-charged	no	3	N-I			Beta sheet	Interface
77	ARG	NDUFS7	Strongly	Positively charged	no	3	R-C			Loop	Interface
81	VAL	NDUFS7	Not	Non-polar	no	3	V-I			Beta sheet	Buried
71	PRO	NDUFS8	Fully	Polar non-charged	no	3	P-S			Loop	Surface
81	LYS	NDUFS8	Fully	Positively charged	no	3	K-E			Loop	Interface
109	TYR	NDUFS8	Fully	Aromatic	no	3	Y-F			Beta sheet	Buried
127	PRO	NDUFS8	Fully	Polar non-charged	no	3	P-T			Loop	Interface
170	GLN	NDUFS8	Fully	Polar non-charged	no	3	Q-L			Helix	Surface
19	ASP	NDUFV1	Fully	Negatively charged	yes	3	D-G			Loop	Buried
75	THR	NDUFV1	Fully	Polar non-charged	no	3	T-A			Helix	Buried
131	ALA	NDUFV1	Not	Non-polar	no	3	A-T			Loop	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
140	GLY	NDUFV1	Weakly	Non-polar	no	3	G-E			Loop	Buried
247	ARG	NDUFV1	Fully	Positively charged	no	3	R-K			Loop	Surface
263	ASN	NDUFV1	Fully	Polar non-charged	no	3	N-S			Loop	Surface
297	VAL	NDUFV1	Fully	Non-polar	no	3	V-A			Beta sheet	Buried
348	ALA	NDUFV1	Not	Non-polar	no	3	A-T			Helix	Buried
153	MET	NDUFV2	Fully	Non-polar	no	3	M-V			Beta sheet	Buried
158	ASP	NDUFV2	Fully	Negatively charged	yes	3	D-G			Loop	Interface
67	THR	ND1	Strongly	Polar non-charged	no			1	T-A, T-I	Loop	Buried
81	ILE	ND1	Strongly	Non-polar	no			1	I-L, I-V, I-T	TMH	Buried
309	ILE	ND1	Not	Non-polar	no			1	I-T, I-V	TMH	Interface
8	VAL	ND2	Not	Non-polar	no			1	V-G, V-I	TMH	Interface
119	THR	ND2	Not	Polar non-charged	no			1	T-A	TMH	Buried
193	VAL	ND2	Not	Non-polar	no			1	V-I	TMH	Buried
237	LEU	ND2	Not	Non-polar	no			1	L-M	Loop	Interface
325	PHE	ND2	Not	Aromatic	no			1	F-L	TMH	Surface
88	VAL	ND3	Not	Non-polar	no			1	V-A, V-I	TMH	Interface
114	THR	ND3	Not	Polar non-charged	no			1	T-A, T-S, T-I	Loop	Surface
140	PRO	ND4	Not	Polar non-charged	no			1	P-S, P-T	Loop	Surface
404	ALA	ND4	Weakly	Non-polar	no			1	A-T, A-V	TMH	Buried
47	MET	ND4L	Not	Non-polar	no			1	M-I, M-T	TMH	Interface
8	THR	ND5	Not	Polar non-charged	no			1	T-A	TMH	Buried
24	VAL	ND5	Not	Non-polar	no			1	V-I	Loop	Surface
257	ILE	ND5	Strongly	Non-polar	no			1	I-V	TMH	Buried
314	MET	ND5	Strongly	Non-polar	no			1	M-V	TMH	Buried
475	ALA	ND5	Not	Non-polar	no			1	A-T	Loop	Surface
482	ILE	ND5	Not	Non-polar	no			1	I-L, I-V, I-T, I-M	Loop	Buried
485	TYR	ND5	Not	Aromatic	no			1	Y-C, Y-H	Loop	Surface
515	SER	ND5	Not	Polar non-charged	no			1	S-F, S-T, S-P, S-Y	Loop	Surface
531	SER	ND5	Not	Polar non-charged	no			1	S-C, S-N, S-T, S-I	Helix	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
533	THR	ND5	Not	Polar non-charged	no			1	T-A, T-M	Helix	Surface
555	LEU	ND5	Not	Non-polar	no			1	L-Q	Helix	Interface
166	ILE	ND6	Strongly	Non-polar	no			1	I-M, I-V	TMH	Interface
47	LYS	NDUFA7	Strongly	Positively charged	no			1	K-Q	Loop	Surface
259	PRO	NDUFA9	Fully	Polar non-charged	no			1	P-L	Loop	Surface
145	PRO	NDUFB9	Not	Polar non-charged	no			1	P-S	Loop	Surface
46	LEU	NDUFC2	Not	Non-polar	no			1	L-V	TMH	Surface
319	PRO	NDUFS2	Fully	Polar non-charged	no			1	P-A, P-L	Loop	Surface
102	VAL	NDUFV2	Strongly	Non-polar	no			1	V-A	Beta sheet	Buried
10	ILE	ND1	Not	Non-polar	no			2	I-T, I-V, I-N	TMH	Buried
29	GLY	ND1	Fully	Non-polar	no			2	G-A, G-S, G-D	TMH	Buried
43	TYR	ND1	Not	Aromatic	no			2	Y-C, Y-H	Loop	Surface
69	THR	ND1	Not	Polar non-charged	no			2	T-A, T-S	TMH	Surface
73	THR	ND1	Not	Polar non-charged	no			2	T-A, T-S	TMH	Surface
87	THR	ND1	Not	Polar non-charged	no			2	T-A	TMH	Buried
89	LEU	ND1	Strongly	Non-polar	no			2	L-F	Loop	Buried
229	THR	ND1	Not	Polar non-charged	no			2	T-A, T-M	TMH	Buried
236	THR	ND1	Not	Polar non-charged	no			2	T-A, T-S, T-I	TMH	Buried
248	ASP	ND1	Not	Negatively charged	no			2	D-N	Loop	Surface
263	THR	ND1	Strongly	Polar non-charged	no			2	T-A, T-I	TMH	Buried
275	THR	ND1	Strongly	Polar non-charged	no			2	T-A	Loop	Interface
307	MET	ND1	Fully	Non-polar	no			2	M-T, M-V	TMH	Buried
13	ILE	ND2	Strongly	Non-polar	no			2	I-V	TMH	Buried
31	VAL	ND2	Not	Non-polar	no			2	V-A, V-M, V-G	TMH	Interface
43	VAL	ND2	Strongly	Non-polar	no			2	V-I	TMH	Buried
76	PHE	ND2	Not	Aromatic	no			2	F-L, F-S	TMH	Buried
80	LEU	ND2	Not	Non-polar	no			2	L-F, L-I	Loop	Surface
88	ASN	ND2	Not	Polar non-charged	no			2	N-S	Loop	Interface
89	THR	ND2	Not	Polar non-charged	no			2	T-A, T-N	Loop	Buried
100	MET	ND2	Not	Non-polar	no			2	M-I, M-V, M-T	TMH	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
152	SER	ND2	Not	Polar non-charged	no			2	S-G, S-N, S-T	TMH	Interface
164	ALA	ND2	Not	Non-polar	no			2	A-T, A-V	TMH	Buried
202	ILE	ND2	Strongly	Non-polar	no			2	I-T, I-V	TMH	Buried
203	LEU	ND2	Strongly	Non-polar	no			2	L-F	TMH	Buried
265	ALA	ND2	Not	Non-polar	no			2	A-T, A-V	TMH	Buried
270	PHE	ND2	Strongly	Aromatic	no			2	F-L	TMH	Buried
278	ILE	ND2	Not	Non-polar	no			2	I-L, I-V, I-T	TMH	Buried
284	THR	ND2	Not	Polar non-charged	no			2	T-A, T-S, T-I	TMH	Surface
288	LEU	ND2	Strongly	Non-polar	no			2	L-I	TMH	Buried
323	THR	ND2	Not	Polar non-charged	no			2	T-A, T-P, T-I	Loop	Surface
342	PHE	ND2	Not	Aromatic	no			2	F-C, F-L	TMH	Interface
9	ILE	ND3	Not	Non-polar	no			2	I-T, I-V	TMH	Interface
29	GLY	ND3	Not	Non-polar	no			2	G-S	Loop	Surface
49	VAL	ND3	Strongly	Non-polar	no			2	V-I	Loop	Surface
96	ILE	ND3	Not	Non-polar	no			2	I-V, I-L, I-T, I-M	TMH	Surface
103	ALA	ND3	Strongly	Non-polar	no			2	A-T, A-V	TMH	Surface
47	ASN	ND4	Not	Polar non-charged	no			2	N-S	Loop	Buried
50	PHE	ND4	Not	Aromatic	no			2	F-L, F-S	Beta sheet	Buried
54	PRO	ND4	Not	Polar non-charged	no			2	P-L	Loop	Surface
86	SER	ND4	Not	Polar non-charged	no			2	S-G, S-N	Loop	Surface
89	LEU	ND4	Not	Non-polar	no			2	L-P	TMH	Surface
101	SER	ND4	Not	Polar non-charged	no			2	S-C, S-F	TMH	Surface
110	PHE	ND4	Fully	Aromatic	no			2	F-L, F-S	TMH	Buried
131	ALA	ND4	Not	Non-polar	no			2	A-T, A-V	TMH	Buried
138	ASN	ND4	Not	Polar non-charged	no			2	N-S	Loop	Surface
149	PHE	ND4	Fully	Aromatic	no			2	F-L	TMH	Buried
230	VAL	ND4	Strongly	Non-polar	no			2	V-L, V-M	TMH	Buried
299	THR	ND4	Not	Polar non-charged	no			2	T-A, T-M	TMH	Buried
390	ASN	ND4	Weakly	Polar non-charged	no			2	N-S	Loop	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
402	VAL	ND4	Strongly	Non-polar	no			2	V-I, V-L	TMH	Buried
424	ASN	ND4	Not	Polar non-charged	no			2	N-S, N-K	Loop	Interface
442	SER	ND4	Not	Polar non-charged	no			2	S-A, S-F	TMH	Buried
445	LEU	ND4	Strongly	Non-polar	no			2	L-F, L-I	TMH	Surface
459	SER	ND4	Not	Polar non-charged	no			2	S-F, S-T, S-P, S-Y	Loop	Interface
13	THR	ND4L	Not	Polar non-charged	no			2	T-A, T-P	TMH	Surface
62	ALA	ND4L	Not	Non-polar	no			2	A-P, A-T, A-V	TMH	Interface
94	ASN	ND4L	Fully	Polar non-charged	no			2	N-S	Loop	Surface
4	HIS	ND5	Not	Positively charged	no			2	H-Y	TMH	Buried
13	THR	ND5	Not	Polar non-charged	no			2	T-A	TMH	Surface
22	THR	ND5	Not	Polar non-charged	no			2	T-A, T-I	TMH	Surface
23	LEU	ND5	Not	Non-polar	no			2	L-F	TMH	Surface
52	THR	ND5	Not	Polar non-charged	no			2	T-C, T-M	Loop	Buried
69	ALA	ND5	Not	Non-polar	no			2	A-T, A-V	Loop	Interface
74	THR	ND5	Not	Polar non-charged	no			2	T-A, T-I	Loop	Buried
182	PHE	ND5	Strongly	Aromatic	no			2	F-L	TMH	Buried
190	ILE	ND5	Not	Non-polar	no			2	I-M, I-V, I-S	TMH	Interface
201	MET	ND5	Strongly	Non-polar	no			2	M-V	Loop	Buried
202	ALA	ND5	Not	Non-polar	no			2	A-T, A-V	Loop	Interface
205	ASN	ND5	Not	Polar non-charged	no			2	N-S, N-T	Loop	Buried
261	ILE	ND5	Strongly	Non-polar	no			2	I-V	TMH	Buried
265	PRO	ND5	Not	Polar non-charged	no			2	P-H, P-S	Loop	Surface
270	SER	ND5	Weakly	Polar non-charged	no			2	S-C, S-N, S-T	Loop	Surface
283	ILE	ND5	Strongly	Non-polar	no			2	I-T, I-V, I-M	TMH	Buried
410	SER	ND5	Not	Polar non-charged	no			2	S-A, S-P, S-F	TMH	Surface
432	THR	ND5	Not	Polar non-charged	no			2	T-A	Loop	Buried
439	THR	ND5	Not	Polar non-charged	no			2	T-A, T-P	Loop	Interface
449	THR	ND5	Not	Polar non-charged	no			2	T-A	TMH	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
459	ALA	ND5	Not	Non-polar	no			2	A-T, A-V	TMH	Surface
469	THR	ND5	Strongly	Polar non-charged	no			2	T-A, T-I	TMH	Surface
471	ASN	ND5	Not	Polar non-charged	no			2	N-S, N-T	TMH	Surface
478	PHE	ND5	Not	Aromatic	no			2	F-I, F-L	Loop	Surface
489	THR	ND5	Not	Polar non-charged	no			2	T-A, T-M	TMH	Surface
495	PHE	ND5	Not	Aromatic	no			2	F-L, F-S	TMH	Surface
518	CYS	ND5	Not	Polar non-charged	no			2	C-R, C-Y, C-S	Helix	Surface
530	PRO	ND5	Not	Polar non-charged	no			2	P-S	Helix	Surface
536	THR	ND5	Not	Polar non-charged	no			2	T-A, T-M	Helix	Buried
538	PRO	ND5	Not	Polar non-charged	no			2	P-S	Helix	Surface
541	GLY	ND5	Not	Non-polar	no			2	G-A	Helix	Surface
549	PRO	ND5	Not	Polar non-charged	no			2	P-S	Helix	Interface
556	THR	ND5	Not	Polar non-charged	no			2	T-A, T-I	Helix	Interface
573	THR	ND5	Not	Polar non-charged	no			2	T-A	Helix	Interface
575	ILE	ND5	Not	Non-polar	no			2	I-M, I-V, I-T	Helix	Surface
592	PHE	ND5	Strongly	Aromatic	no			2	F-L	TMH	Surface
598	THR	ND5	Not	Polar non-charged	no			2	T-A, T-I	TMH	Interface
14	MET	ND6	Not	Non-polar	no			2	M-L, M-V	TMH	Surface
31	VAL	ND6	Not	Non-polar	no			2	V-A	TMH	Buried
41	VAL	ND6	Not	Non-polar	no			2	V-A, V-I	TMH	Surface
86	VAL	ND6	Not	Non-polar	no			2	V-A	Loop	Surface
97	ALA	ND6	Not	Non-polar	no			2	A-P, A-V, A-S	TMH	Surface
120	SER	ND6	Not	Polar non-charged	no			2	S-G, S-M, S-N	Loop	Buried
139	PRO	ND6	Not	Polar non-charged	no			2	P-L, P-S	Helix	Interface
156	THR	ND6	Not	Polar non-charged	no			2	T-A, T-I	TMH	Interface
159	THR	ND6	Not	Polar non-charged	no			2	T-A, T-M, T-P	TMH	Interface
165	TYR	ND6	Not	Aromatic	no			2	Y-H, Y-C, Y-S	TMH	Interface
30	ASN	NDUFA10	Fully	Polar non-charged	no			2	N-S	Beta sheet	Buried
13	LEU	NDUFA2	Not	Non-polar	no			2	L-V	Loop	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
20	VAL	NDUFA3	Weakly	Non-polar	no			2	V-L	TMH	Surface
107	HIS	NDUFA6	Fully	Positively charged	no			2	H-Y, H-Q	Loop	Surface
65	PRO	NDUFA7	Fully	Polar non-charged	no			2	P-A	Loop	Interface
18	MET	NDUFB1	Not	Non-polar	no			2	M-L	TMH	Surface
141	ALA	NDUFB10	Weakly	Non-polar	no			2	A-V, A-T	Loop	Interface
52	PRO	NDUFB2	Not	Polar non-charged	no			2	P-L	Loop	Surface
62	LEU	NDUFB7	Fully	Non-polar	no			2	L-V	Helix	Buried
96	SER	NDUFB8	Weakly	Polar non-charged	no			2	S-T	Loop	Interface
120	VAL	NDUFB8	Not	Non-polar	no			2	V-M	TMH	Surface
8	VAL	NDUFV2	Fully	Non-polar	no			2	V-M	Loop	Interface
5	ASN	ND1	Fully	Polar non-charged	no			3	N-D	TMH	Interface
15	ILE	ND1	Not	Non-polar	no			3	I-T, I-V, I-M	TMH	Surface
17	MET	ND1	Strongly	Non-polar	no			3	M-V, M-I	TMH	Buried
21	MET	ND1	Not	Non-polar	no			3	M-T, M-V	TMH	Buried
27	ILE	ND1	Strongly	Non-polar	no			3	I-F, I-V, I-T, I-M	TMH	Buried
68	ILE	ND1	Not	Non-polar	no			3	I-M, I-V, I-T	TMH	Surface
76	THR	ND1	Not	Polar non-charged	no			3	T-I	TMH	Interface
77	LEU	ND1	Strongly	Non-polar	no			3	L-M	TMH	Surface
96	VAL	ND1	Strongly	Non-polar	no			3	V-A, V-I	Loop	Interface
112	ALA	ND1	Not	Non-polar	no			3	A-T	TMH	Surface
116	ILE	ND1	Fully	Non-polar	no			3	I-N, I-V	TMH	Buried
139	THR	ND1	Fully	Polar non-charged	no			3	T-A	TMH	Buried
152	SER	ND1	Fully	Polar non-charged	no			3	S-A	TMH	Buried
153	THR	ND1	Not	Polar non-charged	no			3	T-A, T-M	TMH	Buried
168	THR	ND1	Fully	Polar non-charged	no			3	T-A	Helix	Interface
171	HIS	ND1	Not	Positively charged	no			3	H-Y	Loop	Surface
196	THR	ND1	Strongly	Polar non-charged	no			3	T-A	Loop	Buried
213	ILE	ND1	Not	Non-polar	no			3	I-M, I-V, I-T	Loop	Interface
239	THR	ND1	Not	Polar non-charged	no			3	T-A, T-S	TMH	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
241	ILE	ND1	Strongly	Non-polar	no			3	I-M	TMH	Buried
249	ALA	ND1	Not	Non-polar	no			3	A-V	Loop	Surface
258	TYR	ND1	Not	Aromatic	no			3	Y-C	TMH	Surface
260	VAL	ND1	Not	Non-polar	no			3	V-I	TMH	Interface
265	LEU	ND1	Not	Non-polar	no			3	L-F	TMH	Surface
266	LEU	ND1	Strongly	Non-polar	no			3	L-M	TMH	Buried
268	SER	ND1	Not	Polar non-charged	no			3	S-F	TMH	Surface
273	ILE	ND1	Strongly	Non-polar	no			3	I-V	TMH	Buried
292	ASN	ND1	Weakly	Polar non-charged	no			3	N-D, N-S	TMH	Surface
300	LEU	ND1	Not	Non-polar	no			3	L-S	TMH	Buried
305	VAL	ND1	Strongly	Non-polar	no			3	V-I	TMH	Interface
310	THR	ND1	Not	Polar non-charged	no			3	T-A, T-S	Loop	Interface
313	SER	ND1	Strongly	Polar non-charged	no			3	S-G	Loop	Interface
318	THR	ND1	Not	Polar non-charged	no			3	T-A, T-I	Loop	Surface
11	SER	ND2	Not	Polar non-charged	no			3	S-F, S-T, S-P	TMH	Surface
15	ALA	ND2	Not	Non-polar	no			3	A-T	TMH	Buried
19	ILE	ND2	Strongly	Non-polar	no			3	I-T	TMH	Buried
48	MET	ND2	Not	Non-polar	no			3	M-T	Loop	Surface
49	ASN	ND2	Not	Polar non-charged	no			3	N-S	Loop	Buried
62	THR	ND2	Not	Polar non-charged	no			3	T-M	TMH	Buried
69	ILE	ND2	Strongly	Non-polar	no			3	I-V	TMH	Buried
78	ASN	ND2	Not	Polar non-charged	no			3	N-S	TMH	Interface
79	MET	ND2	Not	Non-polar	no			3	M-L, M-V, M-T	TMH	Interface
81	SER	ND2	Not	Polar non-charged	no			3	S-F, S-T	Loop	Interface
85	THR	ND2	Not	Polar non-charged	no			3	T-A	Loop	Interface
94	SER	ND2	Not	Polar non-charged	no			3	S-P	TMH	Buried
98	ILE	ND2	Not	Non-polar	no			3	I-M	TMH	Buried
115	VAL	ND2	Strongly	Non-polar	no			3	V-A, V-I, V-L	TMH	Buried
125	THR	ND2	Not	Polar non-charged	no			3	T-M, T-S	TMH	Surface
139	ILE	ND2	Not	Non-polar	no			3	I-L	TMH	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
141	ILE	ND2	Strongly	Non-polar	no			3	I-V	TMH	Buried
145	ILE	ND2	Not	Non-polar	no			3	I-L, I-M	TMH	Buried
147	PRO	ND2	Not	Polar non-charged	no			3	P-S	Loop	Surface
148	SER	ND2	Not	Polar non-charged	no			3	S-A	Loop	Surface
156	THR	ND2	Not	Polar non-charged	no			3	T-I, T-S	TMH	Surface
159	ILE	ND2	Not	Non-polar	no			3	I-V	TMH	Buried
166	SER	ND2	Weakly	Polar non-charged	no			3	S-G, S-N	TMH	Buried
187	MET	ND2	Strongly	Non-polar	no			3	M-T, M-V	TMH	Buried
200	MET	ND2	Not	Non-polar	no			3	M-I, M-T, M-V	TMH	Buried
209	ILE	ND2	Not	Non-polar	no			3	I-T, I-V	TMH	Buried
219	LEU	ND2	Strongly	Non-polar	no			3	L-I, L-V	TMH	Buried
220	ASN	ND2	Not	Polar non-charged	no			3	N-D, N-Y, N-S	TMH	Surface
222	ASN	ND2	Not	Polar non-charged	no			3	N-S	Loop	Buried
238	THR	ND2	Not	Polar non-charged	no			3	T-A	Loop	Interface
239	TRP	ND2	Not	Aromatic	no			3	W-C	Loop	Surface
241	THR	ND2	Not	Polar non-charged	no			3	T-M	TMH	Buried
242	PRO	ND2	Not	Polar non-charged	no			3	P-L	TMH	Surface
246	SER	ND2	Not	Polar non-charged	no			3	S-A, S-P, S-C, S-F	TMH	Buried
266	ILE	ND2	Fully	Non-polar	no			3	I-V	TMH	Buried
267	ILE	ND2	Strongly	Non-polar	no			3	I-V	TMH	Interface
276	LEU	ND2	Not	Non-polar	no			3	L-F, L-I	Loop	Surface
277	ILE	ND2	Not	Non-polar	no			3	I-V	TMH	Surface
281	ILE	ND2	Not	Non-polar	no			3	I-M, I-V	TMH	Surface
285	ILE	ND2	Not	Non-polar	no			3	I-L	TMH	Buried
286	THR	ND2	Strongly	Polar non-charged	no			3	T-S	TMH	Buried
317	PHE	ND2	Not	Aromatic	no			3	F-L, F-V	Loop	Interface
320	THR	ND2	Not	Polar non-charged	no			3	T-A	Loop	Surface
330	ILE	ND2	Not	Non-polar	no			3	I-V	TMH	Buried
332	LEU	ND2	Not	Non-polar	no			3	L-F	TMH	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
333	THR	ND2	Strongly	Polar non-charged	no			3	T-A	TMH	Buried
343	MET	ND2	Not	Non-polar	no			3	M-V	TMH	Surface
345	MET	ND2	Not	Non-polar	no			3	M-I, M-T, M-V	TMH	Interface
346	ILE	ND2	Not	Non-polar	no			3	I-M, I-V	TMH	Surface
14	ALA	ND3	Strongly	Non-polar	no			3	A-S, A-T	TMH	Interface
19	ILE	ND3	Not	Non-polar	no			3	I-T	TMH	Surface
24	LEU	ND3	Fully	Non-polar	no			3	L-M	Loop	Surface
44	MET	ND3	Not	Non-polar	no			3	M-V	Loop	Interface
89	MET	ND3	Not	Non-polar	no			3	M-T, M-V	TMH	Surface
90	SER	ND3	Not	Polar non-charged	no			3	S-L, S-P, S-T	TMH	Interface
93	LEU	ND3	Not	Non-polar	no			3	L-S	TMH	Surface
97	ILE	ND3	Strongly	Non-polar	no			3	I-T	TMH	Interface
105	GLU	ND3	Fully	Negatively charged	no			3	E-D, E-K	TMH	Buried
6	VAL	ND4	Not	Non-polar	no			3	V-A	TMH	Interface
21	HIS	ND4	Not	Positively charged	no			3	H-Q	Loop	Surface
29	THR	ND4	Not	Polar non-charged	no			3	T-I	TMH	Interface
45	ILE	ND4	Not	Non-polar	no			3	I-F	Loop	Interface
46	ASN	ND4	Not	Polar non-charged	no			3	N-D	Loop	Interface
52	CYS	ND4	Not	Polar non-charged	no			3	C-Y	Beta sheet	Surface
55	THR	ND4	Not	Polar non-charged	no			3	T-A, T-I	Loop	Surface
58	SER	ND4	Not	Polar non-charged	no			3	S-P	Beta sheet	Buried
70	THR	ND4	Not	Polar non-charged	no			3	T-A	TMH	Buried
85	SER	ND4	Not	Polar non-charged	no			3	S-F, S-Y	Loop	Surface
104	ILE	ND4	Not	Non-polar	no			3	I-V	TMH	Buried
107	ILE	ND4	Fully	Non-polar	no			3	I-V	TMH	Buried
121	PHE	ND4	Strongly	Aromatic	no			3	F-L	TMH	Buried
129	THR	ND4	Fully	Polar non-charged	no			3	T-S	TMH	Buried
166	TYR	ND4	Not	Aromatic	no			3	Y-H	TMH	Interface
170	THR	ND4	Not	Polar non-charged	no			3	T-I	TMH	Surface
187	SER	ND4	Not	Polar non-charged	no			3	S-A, S-P, S-T	Loop	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
193	ASN	ND4	Not	Polar non-charged	no			3	N-S	TMH	Buried
201	MET	ND4	Not	Non-polar	no			3	M-T	TMH	Surface
225	ILE	ND4	Fully	Non-polar	no			3	I-V	Loop	Buried
232	ALA	ND4	Fully	Non-polar	no			3	A-T	TMH	Buried
291	ILE	ND4	Strongly	Non-polar	no			3	I-V	TMH	Buried
317	ILE	ND4	Strongly	Non-polar	no			3	I-V	TMH	Buried
320	GLY	ND4	Fully	Non-polar	no			3	G-E	TMH	Buried
350	THR	ND4	Not	Polar non-charged	no			3	T-A	Loop	Interface
365	ALA	ND4	Strongly	Non-polar	no			3	A-T	TMH	Buried
391	ILE	ND4	Not	Non-polar	no			3	I-T, I-V	Loop	Surface
418	SER	ND4	Not	Polar non-charged	no			3	S-A, S-P, S-L	Loop	Interface
419	LEU	ND4	Not	Non-polar	no			3	L-F, L-P	Loop	Buried
420	THR	ND4	Not	Polar non-charged	no			3	T-A	Loop	Buried
421	HIS	ND4	Not	Positively charged	no			3	H-Y, H-P	Loop	Surface
425	ASN	ND4	Not	Polar non-charged	no			3	N-S	Loop	Interface
434	ASN	ND4	Not	Polar non-charged	no			3	N-H	TMH	Buried
435	THR	ND4	Not	Polar non-charged	no			3	T-I	TMH	Interface
455	THR	ND4	Not	Polar non-charged	no			3	T-I	Loop	Interface
21	VAL	ND4L	Not	Non-polar	no			3	V-L, V-M	Loop	Surface
59	VAL	ND4L	Not	Non-polar	no			3	V-M, V-G	TMH	Interface
79	VAL	ND4L	Not	Non-polar	no			3	V-I	TMH	Interface
80	SER	ND4L	Not	Polar non-charged	no			3	S-A	TMH	Interface
2	THR	ND5	Not	Polar non-charged	no			3	T-A, T-I	TMH	Surface
14	SER	ND5	Not	Polar non-charged	no			3	S-A	TMH	Buried
18	PRO	ND5	Fully	Polar non-charged	no			3	P-L	TMH	Buried
27	ASN	ND5	Not	Polar non-charged	no			3	N-S	Loop	Surface
30	ASN	ND5	Not	Polar non-charged	no			3	N-S	Loop	Surface
34	HIS	ND5	Not	Positively charged	no			3	H-R, H-Y	TMH	Surface
38	SER	ND5	Not	Polar non-charged	no			3	S-T	TMH	Surface
39	ILE	ND5	Not	Non-polar	no			3	I-V	TMH	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
40	VAL	ND5	Not	Non-polar	no			3	V-I	TMH	Buried
45	ILE	ND5	Not	Non-polar	no			3	I-V	TMH	Surface
53	MET	ND5	Not	Non-polar	no			3	M-V	Loop	Surface
59	GLN	ND5	Not	Polar non-charged	no			3	Q-L	Loop	Interface
63	ILE	ND5	Weakly	Non-polar	no			3	I-V	Beta sheet	Buried
65	ASN	ND5	Weakly	Polar non-charged	no			3	N-S	Beta sheet	Surface
67	HIS	ND5	Not	Positively charged	no			3	H-Y	Loop	Surface
70	THR	ND5	Not	Polar non-charged	no			3	T-A, T-M	Loop	Interface
76	LEU	ND5	Strongly	Non-polar	no			3	L-F, L-V, L-H	Loop	Buried
88	MET	ND5	Not	Non-polar	no			3	M-T	TMH	Buried
93	ALA	ND5	Fully	Non-polar	no			3	A-S, A-T	TMH	Buried
109	ASN	ND5	Not	Polar non-charged	no			3	N-D, N-K, N-Y, N-S	Loop	Surface
113	ASN	ND5	Not	Polar non-charged	no			3	N-S	Loop	Surface
127	THR	ND5	Strongly	Polar non-charged	no			3	T-A, T-I	TMH	Buried
132	VAL	ND5	Not	Non-polar	no			3	V-F, V-I	TMH	Buried
141	PHE	ND5	Strongly	Aromatic	no			3	F-L, F-C	TMH	Buried
147	VAL	ND5	Strongly	Non-polar	no			3	V-M	TMH	Buried
162	ALA	ND5	Strongly	Non-polar	no			3	A-T	TMH	Interface
169	ILE	ND5	Strongly	Non-polar	no			3	I-T, I-V	TMH	Interface
172	ILE	ND5	Strongly	Non-polar	no			3	I-V	TMH	Buried
177	ILE	ND5	Strongly	Non-polar	no			3	I-M, I-V	TMH	Surface
196	TRP	ND5	Fully	Aromatic	no			3	W-G, W-L	Loop	Buried
206	ALA	ND5	Not	Non-polar	no			3	A-T	Loop	Surface
209	SER	ND5	Not	Polar non-charged	no			3	S-G	Loop	Buried
211	THR	ND5	Not	Polar non-charged	no			3	T-A, T-P	Loop	Surface
271	PRO	ND5	Not	Polar non-charged	no			3	P-L, P-S	TMH	Surface
273	ILE	ND5	Not	Non-polar	no			3	I-T, I-V	TMH	Buried
290	VAL	ND5	Strongly	Non-polar	no			3	V-I	TMH	Surface
304	PHE	ND5	Fully	Aromatic	no			3	F-L	TMH	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
317	ILE	ND5	Strongly	Non-polar	no			3	I-V	TMH	Buried
323	HIS	ND5	Strongly	Positively charged	no			3	H-Y	TMH	Buried
331	THR	ND5	Not	Polar non-charged	no			3	T-A	TMH	Buried
377	SER	ND5	Weakly	Polar non-charged	no			3	S-N, S-T	TMH	Buried
378	LEU	ND5	Strongly	Non-polar	no			3	L-P, L-Q	TMH	Interface
381	ALA	ND5	Not	Non-polar	no			3	A-P, A-V, A-T	TMH	Surface
400	ASN	ND5	Weakly	Polar non-charged	no			3	N-S	Loop	Interface
403	TYR	ND5	Not	Aromatic	no			3	Y-D, Y-N, Y-H	Loop	Surface
413	LEU	ND5	Fully	Non-polar	no			3	L-I	TMH	Surface
420	SER	ND5	Strongly	Polar non-charged	no			3	S-G	TMH	Buried
423	SER	ND5	Fully	Polar non-charged	no			3	S-R, S-T	TMH	Buried
424	THR	ND5	Not	Polar non-charged	no			3	T-A	TMH	Buried
440	LEU	ND5	Not	Non-polar	no			3	L-F	Loop	Surface
441	THR	ND5	Not	Polar non-charged	no			3	T-I	Loop	Interface
442	ASN	ND5	Not	Polar non-charged	no			3	N-D, N-S	Loop	Surface
447	ASN	ND5	Not	Polar non-charged	no			3	N-D, N-S	Loop	Surface
456	ARG	ND5	Fully	Positively charged	no			3	R-G	TMH	Surface
461	SER	ND5	Fully	Polar non-charged	no			3	S-T, S-G	TMH	Buried
472	ILE	ND5	Strongly	Non-polar	no			3	I-M	Loop	Buried
473	SER	ND5	Not	Polar non-charged	no			3	S-F, S-P, S-T	Loop	Surface
476	SER	ND5	Not	Polar non-charged	no			3	S-A, S-T, S-F	Loop	Interface
477	PRO	ND5	Not	Polar non-charged	no			3	P-H, P-S, P-L	Loop	Interface
491	LEU	ND5	Strongly	Non-polar	no			3	L-F	TMH	Buried
492	ALA	ND5	Not	Non-polar	no			3	A-T	TMH	Surface
493	VAL	ND5	Not	Non-polar	no			3	V-I	TMH	Surface
500	THR	ND5	Not	Polar non-charged	no			3	T-A, T-S, T-M	TMH	Surface
508	THR	ND5	Not	Polar non-charged	no			3	T-A	TMH	Surface
509	ASN	ND5	Not	Polar non-charged	no			3	N-S	Loop	Surface
513	MET	ND5	Not	Non-polar	no			3	M-T	Loop	Surface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
517	LEU	ND5	Not	Non-polar	no			3	L-M, L-V	Loop	Surface
524	ASN	ND5	Weakly	Polar non-charged	no			3	N-S	Helix	Interface
525	MET	ND5	Not	Non-polar	no			3	M-L	Loop	Surface
546	GLN	ND5	Not	Polar non-charged	no			3	Q-L	Helix	Interface
564	LYS	ND5	Fully	Positively charged	no			3	K-M	Helix	Surface
565	THR	ND5	Not	Polar non-charged	no			3	T-A, T-M	Helix	Surface
569	HIS	ND5	Not	Positively charged	no			3	H-Y	Helix	Surface
571	ILE	ND5	Not	Non-polar	no			3	I-T, I-V	Helix	Interface
577	THR	ND5	Not	Polar non-charged	no			3	T-A, T-I	Helix	Surface
579	THR	ND5	Weakly	Polar non-charged	no			3	T-A	Loop	Interface
596	ILE	ND5	Not	Non-polar	no			3	I-L, I-V	TMH	Surface
602	ILE	ND5	Not	Non-polar	no			3	I-M, I-S	Loop	Surface
603	THR	ND5	Not	Polar non-charged	no			3	T-M	Loop	Surface
2	MET	ND6	Not	Non-polar	no			3	M-I, M-T	Loop	Surface
4	ALA	ND6	Not	Non-polar	no			3	A-T, A-V	TMH	Surface
34	VAL	ND6	Not	Non-polar	no			3	V-A, V-I	TMH	Surface
35	SER	ND6	Strongly	Polar non-charged	no			3	S-C	TMH	Buried
37	VAL	ND6	Not	Non-polar	no			3	V-A	TMH	Surface
38	VAL	ND6	Not	Non-polar	no			3	V-I, V-L	TMH	Buried
42	ILE	ND6	Strongly	Non-polar	no			3	I-V	TMH	Buried
45	ASN	ND6	Not	Polar non-charged	no			3	N-S	TMH	Interface
54	MET	ND6	Strongly	Non-polar	no			3	M-I, M-V	TMH	Interface
81	ALA	ND6	Not	Non-polar	no			3	A-T	Loop	Buried
85	GLY	ND6	Not	Non-polar	no			3	G-E	Loop	Surface
90	VAL	ND6	Not	Non-polar	no			3	V-A	TMH	Surface
94	VAL	ND6	Not	Non-polar	no			3	V-A, V-M	TMH	Surface
102	LEU	ND6	Not	Non-polar	no			3	L-F	TMH	Surface
106	VAL	ND6	Not	Non-polar	no			3	V-A	TMH	Surface
121	VAL	ND6	Not	Non-polar	no			3	V-A, V-M	Loop	Buried
134	LEU	ND6	Not	Non-polar	no			3	L-F, L-M	Loop	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
135	ILE	ND6	Weakly	Non-polar	no			3	I-V	Loop	Interface
140	ILE	ND6	Not	Non-polar	no			3	I-V	Helix	Interface
142	ALA	ND6	Not	Non-polar	no			3	A-V	Helix	Interface
150	ARG	ND6	Not	Positively charged	no			3	R-C, R-P	TMH	Surface
155	VAL	ND6	Strongly	Non-polar	no			3	V-A, V-I	TMH	Interface
161	PHE	ND6	Strongly	Aromatic	no			3	F-L	TMH	Interface
162	VAL	ND6	Not	Non-polar	no			3	V-A	TMH	Interface
174	ASN	ND6	Not	Polar non-charged	no			3	N-S, N-H	Loop	Surface
2	ARG	NDUFA10	Not	Positively charged	no			3	R-C	Loop	Surface
137	ALA	NDUFA10	Fully	Non-polar	no			3	A-V	Helix	Buried
142	GLY	NDUFA10	Not	Non-polar	no			3	G-R	Loop	Surface
165	PRO	NDUFA10	Fully	Polar non-charged	no			3	P-L	Loop	Buried
182	ARG	NDUFA10	Strongly	Positively charged	no			3	R-W, R-Q	Helix	Buried
19	ARG	NDUFA11	Fully	Positively charged	no			3	R-H	TMH	Surface
65	VAL	NDUFA11	Not	Non-polar	no			3	V-F	TMH	Buried
99	THR	NDUFA11	Not	Polar non-charged	no			3	T-I, T-S	TMH	Buried
111	ALA	NDUFA11	Weakly	Non-polar	no			3	A-T, A-V	TMH	Surface
128	ARG	NDUFA11	Strongly	Positively charged	no			3	R-W, R-Q	TMH	Surface
60	ARG	NDUFA12	Fully	Positively charged	yes			3	R-Q, R-G	Loop	Buried
9	MET	NDUFA13	Fully	Non-polar	no			3	M-T, M-I	Loop	Surface
24	LEU	NDUFA13	Fully	Non-polar	no			3	L-W, L-F	Loop	Surface
41	THR	NDUFA13	Strongly	Polar non-charged	no			3	T-N, T-I	TMH	Surface
69	ALA	NDUFA13	Fully	Non-polar	no			3	A-T, A-V, A-S	TMH	Interface
80	ARG	NDUFA13	Fully	Positively charged	yes			3	R-Q	TMH	Interface
87	ARG	NDUFA13	Fully	Positively charged	no			3	R-W	TMH	Interface
34	ASP	NDUFA2	Not	Negatively charged	no			3	D-N, D-E	Helix	Surface
62	PRO	NDUFA2	Not	Polar non-charged	no			3	P-L	Loop	Buried
63	LYS	NDUFA2	Strongly	Positively charged	no			3	K-R	Beta sheet	Buried
93	VAL	NDUFA2	Fully	Non-polar	no			3	V-I	Helix	Buried
11	ALA	NDUFA3	Weakly	Non-polar	no			3	A-G	Helix	Buried

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
49	PRO	NDUFA3	Fully	Polar non-charged	no			3	P-T, P-R	Loop	Interface
2	GLY	NDUFA5	Weakly	Non-polar	no			3	G-C	Loop	Surface
63	ARG	NDUFA6	Fully	Positively charged	yes			3	R-Q	Helix	Surface
4	THR	NDUFA7	Fully	Polar non-charged	no			3	T-P, T-A	Loop	Surface
7	ILE	NDUFA7	Strongly	Non-polar	no			3	I-V, I-S	Helix	Buried
29	GLU	NDUFA7	Fully	Negatively charged	no			3	E-K	Loop	Surface
37	PRO	NDUFA7	Not	Polar non-charged	no			3	P-S	Loop	Surface
50	ASN	NDUFA7	Not	Polar non-charged	no			3	N-S	Loop	Surface
57	ASP	NDUFA7	Fully	Negatively charged	no			3	D-G	Loop	Buried
59	ARG	NDUFA7	Fully	Positively charged	no			3	R-C	Loop	Interface
43	MET	NDUFA8	Fully	Non-polar	no			3	M-V, M-I	Helix	Interface
54	ARG	NDUFA8	Strongly	Positively charged	yes			3	R-W, R-Q	Loop	Surface
134	ARG	NDUFA8	Fully	Positively charged	yes			3	R-Q	Loop	Surface
21	ALA	NDUFA9	Fully	Non-polar	no			3	A-V	Beta sheet	Buried
135	LEU	NDUFA9	Fully	Non-polar	no			3	L-M	Loop	Buried
187	ARG	NDUFA9	Not	Positively charged	no			3	R-Q, R-W	Loop	Surface
206	TYR	NDUFA9	Fully	Aromatic	no			3	Y-C, Y-H	Loop	Buried
235	ARG	NDUFA9	Strongly	Positively charged	yes			3	R-Q, R-W	Loop	Buried
249	ALA	NDUFA9	Strongly	Non-polar	no			3	A-D, A-T	Loop	Buried
342	ILE	NDUFA9	Not	Non-polar	no			3	I-M, I-T	Loop	Surface
33	ARG	NDUFB1	Strongly	Positively charged	no			3	R-W, R-Q, R-L	TMH	Surface
38	ARG	NDUFB1	Not	Positively charged	no			3	R-W, R-Q, R-P	Loop	Buried
34	VAL	NDUFB10	Strongly	Non-polar	no			3	V-L	Helix	Interface
65	PRO	NDUFB10	Fully	Polar non-charged	no			3	P-R	Loop	Surface
109	ARG	NDUFB10	Fully	Positively charged	no			3	R-M, R-S	Helix	Interface
139	LEU	NDUFB10	Fully	Non-polar	no			3	L-M, L-P	Loop	Buried
164	LYS	NDUFB10	Strongly	Positively charged	no			3	K-E	Loop	Surface
55	VAL	NDUFB11	Strongly	Non-polar	no			3	V-I	TMH	Surface
56	TRP	NDUFB11	Fully	Aromatic	no			3	W-C	TMH	Surface
10	ARG	NDUFB2	Fully	Positively charged	no			3	R-W, R-Q	Loop	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
60	VAL	NDUFB3	Not	Non-polar	no			3	V-A	Loop	Interface
2	PHE	NDUFB4	Not	Aromatic	no			3	F-L	Loop	Surface
25	SER	NDUFB4	Fully	Polar non-charged	no			3	S-F	Loop	Surface
66	ARG	NDUFB4	Not	Positively charged	no			3	R-C, R-H, R-L	Helix	Interface
11	ILE	NDUFB5	Strongly	Non-polar	no			3	I-F, I-M	Loop	Surface
87	TYR	NDUFB5	Fully	Aromatic	no			3	Y-H	Helix	Interface
137	PRO	NDUFB5	Not	Polar non-charged	no			3	P-L	Loop	Surface
10	ARG	NDUFB6	Fully	Positively charged	no			3	R-Q	Loop	Interface
20	ARG	NDUFB6	Not	Positively charged	no			3	R-Q	Loop	Surface
25	GLN	NDUFB6	Fully	Polar non-charged	no			3	Q-E	Loop	Interface
42	MET	NDUFB6	Not	Non-polar	no			3	M-V	Loop	Surface
80	ILE	NDUFB6	Strongly	Non-polar	no			3	I-T	TMH	Surface
108	THR	NDUFB6	Not	Polar non-charged	no			3	T-A, T-I	Loop	Surface
29	GLY	NDUFB7	Fully	Non-polar	no			3	G-C	Loop	Surface
43	GLN	NDUFB7	Fully	Polar non-charged	no			3	Q-R	Loop	Surface
48	ALA	NDUFB7	Fully	Non-polar	no			3	A-T, A-V	Loop	Interface
55	ARG	NDUFB7	Fully	Positively charged	no			3	R-Q	Loop	Interface
61	HIS	NDUFB7	Strongly	Positively charged	no			3	H-D	Helix	Interface
70	ARG	NDUFB7	Fully	Positively charged	yes			3	R-C	Helix	Interface
83	ARG	NDUFB7	Strongly	Positively charged	yes			3	R-W	Helix	Surface
92	ARG	NDUFB7	Not	Positively charged	no			3	R-H	Helix	Surface
105	ARG	NDUFB7	Fully	Positively charged	no			3	R-G, R-Q	Helix	Surface
110	ARG	NDUFB7	Strongly	Positively charged	yes			3	R-Q	Helix	Interface
112	LYS	NDUFB7	Strongly	Positively charged	no			3	K-Q	Helix	Surface
28	MET	NDUFB8	Strongly	Non-polar	no			3	M-I	Loop	Buried
38	ASP	NDUFB8	Fully	Negatively charged	no			3	D-N	Loop	Surface
75	PRO	NDUFB8	Not	Polar non-charged	no			3	P-L	Loop	Buried
87	ARG	NDUFB8	Not	Positively charged	no			3	R-C, R-H	Loop	Interface
93	THR	NDUFB8	Not	Polar non-charged	no			3	T-A	Loop	Surface
94	PRO	NDUFB8	Not	Polar non-charged	no			3	P-R	Loop	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
3	LEU	NDUFB9	Not	Non-polar	no			3	L-W	Loop	Surface
7	PRO	NDUFB9	Not	Polar non-charged	no			3	P-L, P-H	Loop	Surface
36	TYR	NDUFB9	Fully	Aromatic	no			3	Y-H	Helix	Buried
46	ARG	NDUFB9	Fully	Positively charged	no			3	R-L	Helix	Surface
128	ARG	NDUFB9	Strongly	Positively charged	no			3	R-Q	Helix	Buried
143	THR	NDUFB9	Fully	Polar non-charged	no			3	T-M	Loop	Surface
174	ARG	NDUFB9	Fully	Positively charged	no			3	R-W, R-Q	Loop	Interface
176	ARG	NDUFB9	Fully	Positively charged	no			3	R-Q	Loop	Surface
178	MET	NDUFB9	Not	Non-polar	no			3	M-T	Loop	Buried
41	GLU	NDUFC1	Fully	Negatively charged	yes			3	E-Q	TMH	Buried
35	PHE	NDUFC2	Strongly	Aromatic	no			3	F-L	TMH	Surface
59	ARG	NDUFC2	Fully	Positively charged	no			3	R-C, R-H	TMH	Buried
119	ARG	NDUFC2	Strongly	Positively charged	yes			3	R-H, R-C	Loop	Interface
6	SER	NDUFS1	Fully	Polar non-charged	no			3	S-G	Loop	Surface
122	MET	NDUFS1	Fully	Non-polar	no			3	M-L	Helix	Interface
280	THR	NDUFS1	Not	Polar non-charged	no			3	T-S	Loop	Buried
331	LEU	NDUFS1	Strongly	Non-polar	no			3	L-H	Helix	Buried
334	LEU	NDUFS1	Fully	Non-polar	no			3	L-F	Helix	Buried
479	MET	NDUFS1	Not	Non-polar	no			3	M-R	Helix	Surface
515	ARG	NDUFS1	Not	Positively charged	yes			3	R-G	Helix	Surface
634	ASP	NDUFS1	Fully	Negatively charged	yes			3	D-N	Loop	Surface
646	ASN	NDUFS1	Weakly	Polar non-charged	no			3	N-S, N-D	Helix	Buried
689	LYS	NDUFS1	Fully	Positively charged	no			3	K-R, K-T	Helix	Surface
80	ILE	NDUFS2	Fully	Non-polar	no			3	I-V	Loop	Interface
379	VAL	NDUFS2	Fully	Non-polar	no			3	V-L	Beta sheet	Buried
409	HIS	NDUFS2	Fully	Positively charged	yes			3	H-Y	Loop	Surface
100	ARG	NDUFS3	Fully	Positively charged	yes			3	R-C	Loop	Interface
94	SER	NDUFS4	Not	Polar non-charged	no			3	S-A	Helix	Surface
115	SER	NDUFS4	Fully	Polar non-charged	no			3	S-T	Loop	Surface
71	THR	NDUFS5	Weakly	Polar non-charged	no			3	T-M	Helix	Interface

Residue number	Identity	Subunit	Conservation	Amino acid type	Salt bridge	Mutation class	Residue change	Polymorphism class	Residue change	Secondary structure	Residue environ.
92	THR	NDUFS5	Fully	Polar non-charged	no			3	T-A	Loop	Surface
97	HIS	NDUFS5	Fully	Positively charged	yes			3	H-Q, H-P, H-Y	Loop	Interface
133	ASP	NDUFS7	Fully	Negatively charged	yes			3	D-N, D-E	Loop	Interface
134	ARG	NDUFS7	Fully	Positively charged	yes			3	R-H, R-L	Loop	Interface
148	THR	NDUFS7	Fully	Polar non-charged	no			3	T-M	Loop	Interface
11	SER	NDUFV1	Strongly	Polar non-charged	no			3	S-P	Loop	Surface
32	ARG	NDUFV1	Fully	Positively charged	no			3	R-S	Loop	Surface
168	GLY	NDUFV1	Not	Non-polar	no			3	G-D	Loop	Surface
175	VAL	NDUFV1	Fully	Non-polar	no			3	V-M	Beta sheet	Buried
339	ARG	NDUFV1	Not	Positively charged	no			3	R-C, R-H	Loop	Surface
376	MET	NDUFV1	Fully	Non-polar	no			3	M-I, M-T	Helix	Buried
385	ARG	NDUFV1	Strongly	Positively charged	no			3	R-L, R-Q, R-W	Loop	Surface
418	LEU	NDUFV1	Fully	Non-polar	no			3	L-M	Helix	Buried
19	THR	NDUFV2	Not	Polar non-charged	no			3	T-N, T-I	Loop	Surface
76	PRO	NDUFV2	Fully	Polar non-charged	no			3	P-T, P-R	Loop	Surface
140	ILE	NDUFV2	Fully	Non-polar	no			3	I-T, I-M	Beta sheet	Surface

Appendix 7.9C Distances of variants from nearest key feature in each grouping (A, B & C). Residue number refers to the mature protein.

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
289	LEU	ND1	1	2	ND1 Glu143	9.44	ND2 Glu34	29.78	NDUFA8 Cys87&Cys99	47.00
45	SER	ND3	1	2	ND3-Cys39	13.28	ND2 Glu34	41.43	NDUFAB1 (α)-acyl chain	18.83
340	ARG	ND4	1	2	ND4L Glu70	69.34	ND4 His220	11.95	NDUFA11 Cys94&Cys114	41.76
64	MET	ND6	1	2	ND6 pi-bulge (Gly61-62)	4.97	ND2 Glu34	20.80	NDUFS5 Cys32&Cys65	40.25
52	ALA	ND1	1	3	Q-site	4.40	ND2 Glu34	51.87	NDUFA9 NADPH	27.86
110	SER	ND1	1	3	ND1 Glu143	3.30	ND2 Glu34	25.37	NDUFS5 Cys32&Cys65	38.98
132	ALA	ND1	1	3	ND1 Glu227	11.54	ND2 Glu34	31.91	NDUFAB1 (α)-acyl chain	36.34
34	SER	ND3	1	3	ND3-Cys39	13.76	ND2 Glu34	47.59	NDUFA9 NADPH	22.99
47	ALA	ND3	1	3	ND3-Cys39	14.09	ND2 Glu34	37.03	NDUFAB1 (α)-acyl chain	23.97
36	GLY	ND6	1	3	ND4L Glu34	10.50	ND2 Glu34	19.99	NDUFS5 Cys32&Cys65	29.86
72	ALA	ND6	1	3	ND4L Glu34	15.31	ND2 Glu34	26.46	NDUFAB1 (α)-acyl chain	33.84
259	MET	NDUFS2	1	3	NDUFS2 H59	16.05	ND2 Glu34	59.36	NDUFAB1 (α)-acyl chain	41.04
393	ALA	NDUFS2	1	3	Q-site	12.70	ND2 Glu34	66.99	NDUFAB1 (α)-acyl chain	33.29
24	GLU	ND1	1		Q-site	7.98	ND2 Glu34	50.87	NDUFA9 NADPH	41.83
131	GLY	ND1	1		ND1 Glu227	11.63	ND2 Glu34	34.94	NDUFAB1 (α)-acyl chain	35.28
143	GLU	ND1	1		ND1 Glu143	0.00	ND2 Glu34	25.13	NDUFS5 Cys32&Cys65	41.32
195	ARG	ND1	1		ND1 Glu227	6.78	ND2 Glu34	41.27	NDUFA9 NADPH	43.77
65	VAL	ND4L	1		ND3 Glu68	8.41	ND2 Glu34	7.99	NDUFS5 Cys32&Cys65	28.60
124	PHE	ND5	1		ND4L Glu70	96.76	ND5 Lys223	10.07	ND5 Cys279&NDUFB8 Cys115	31.97
393	ASP	ND5	1		ND4L Glu70	119.39	ND5 Lys392	3.78	ND5 Cys279&NDUFB8 Cys115	19.22
60	LEU	ND6	1		ND6 pi-bulge (Gly61-62)	3.76	ND2 Glu34	19.62	NDUFS5 Cys32&Cys65	34.53
63	MET	ND6	1		ND6 pi-bulge (Gly61-62)	3.78	ND2 Glu34	18.25	NDUFS5 Cys32&Cys65	37.29
8	GLY	NDUFA1	1		Q-site	13.60	ND2 Glu34	56.24	NDUFA8 Cys87&Cys99	40.43
37	ARG	NDUFA1	1		ND1 Glu192	30.76	ND2 Glu34	44.18	NDUFA8 Cys87&Cys99	13.59
107	GLN	NDUFA10	1		ND4L Glu70	46.90	ND2 Leu131 pi bulge	36.61	NDUFA10 ADP	13.21
286	ARG	NDUFA9	1		ND3-Cys39	43.71	ND2 TMH4 Gly107-Phe111	65.50	NDUFA9 NADPH	9.34
208	LEU	NDUFS1	1		Fe-S cluster	6.15	ND2 Glu34	114.83	NDUFS6 Zinc	21.26
218	ARG	NDUFS1	1		Fe-S cluster	10.89	ND2 Glu34	99.45	NDUFS6 Zinc	22.76
229	ASP	NDUFS1	1		Fe-S cluster	25.80	ND2 Glu34	109.91	NDUFS6 Zinc	42.69
385	ARG	NDUFS1	1		Fe-S cluster	23.56	ND2 Glu34	126.83	NDUFS6 Zinc	44.46
572	THR	NDUFS1	1		Fe-S cluster	35.59	ND2 TMH4 Gly107-Phe111	104.15	NDUFAB1 (α)-acyl chain	39.56
596	ASP	NDUFS1	1		Fe-S cluster	40.10	ND2 TMH4 Gly107-Phe111	100.58	NDUFAB1 (α)-acyl chain	39.58
195	ARG	NDUFS2	1		Fe-S cluster	15.28	ND2 Glu34	84.17	NDUFS6 Zinc	33.40
196	PRO	NDUFS2	1		Fe-S cluster	17.84	ND2 Glu34	85.86	NDUFS6 Zinc	34.58
380	SER	NDUFS2	1		Fe-S cluster	24.19	ND2 Glu34	86.19	NDUFAB1 (α)-acyl chain	38.72
109	THR	NDUFS3	1		Q-site	32.69	ND2 Leu131 pi bulge	78.68	NDUFA6 LYR (34-36)	30.86
163	ARG	NDUFS3	1		Fe-S cluster	24.42	ND2 Glu34	76.77	NDUFAB1 (α)-acyl chain	21.39
87	CYS	NDUFS6	1		Fe-S cluster	11.70	ND2 Glu34	107.75	NDUFS6 Zinc	3.82

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
84	VAL	NDUFS7	1		Fe-S cluster	9.80	ND2 Glu34	62.65	NDUFA9 NADPH	28.58
29	GLU	NDUFS8	1		NDUFS2 D160	22.34	ND2 Glu34	50.06	NDUFA8 Cys87&Cys99	50.03
43	ARG	NDUFS8	1		Q-site	22.28	ND2 Glu34	65.11	NDUFA8 Cys87&Cys99	52.11
45	PRO	NDUFS8	1		Q-site	18.89	ND2 Glu34	65.23	NDUFA9 NADPH	48.29
51	PRO	NDUFS8	1		Fe-S cluster	23.74	ND2 Glu34	73.96	NDUFA9 NADPH	36.84
68	ARG	NDUFS8	1		Fe-S cluster	9.11	ND2 Glu34	95.16	NDUFS6 Zinc	15.97
104	ARG	NDUFS8	1		Fe-S cluster	10.51	ND2 Glu34	105.38	NDUFS6 Zinc	13.01
125	ALA	NDUFS8	1		Fe-S cluster	6.14	ND2 Glu34	92.41	NDUFS6 Zinc	19.82
97	ALA	NDUFV1	1		FMN	6.03	ND2 Glu34	148.68	NDUFS6 Zinc	50.70
102	PRO	NDUFV1	1		Fe-S cluster	4.98	ND2 Glu34	143.08	NDUFS6 Zinc	44.12
186	CYS	NDUFV1	1		FMN	6.60	ND2 Glu34	139.85	NDUFS6 Zinc	42.00
194	GLU	NDUFV1	1		Fe-S cluster	11.96	ND2 Glu34	138.64	NDUFS6 Zinc	43.17
403	THR	NDUFV1	1		Fe-S cluster	4.64	ND2 Glu34	125.44	NDUFS6 Zinc	36.23
30	TYR	ND1	2	1	Q-site	14.55	ND2 Glu34	59.27	NDUFA9 NADPH	47.00
164	THR	ND1	2	1	ND1 Glu192	21.53	ND2 Glu34	37.63	NDUFA8 Cys87&Cys99	18.49
240	THR	ND1	2	1	ND1 Glu192	14.93	ND2 Glu34	40.19	NDUFA8 Cys87&Cys99	24.70
1	MET	ND1	2	2	Q-site	27.60	ND2 Glu34	48.74	NDUFA8 Cys87&Cys99	28.00
31	MET	ND1	2	2	Q-site	13.90	ND2 Glu34	57.03	NDUFA9 NADPH	48.76
277	TYR	ND1	2	2	NDUFS2 D160	12.27	ND2 Glu34	46.74	NDUFA8 Cys87&Cys99	51.03
57	ILE	ND2	2	2	ND4L Glu70	12.56	ND2 Lys105	14.14	NDUFA6 LYR (34-36)	42.25
60	ILE	ND3	2	2	ND3 Asp66	8.68	ND2 Glu34	17.03	NDUFS5 Cys32&Cys65	41.75
109	THR	ND4	2	2	ND4L Glu70	55.19	ND4 Glu123	9.52	NDUFB10 Cys70&Cys77	29.55
165	ILE	ND4	2	2	ND4L Glu70	44.55	ND2 Lys263	12.44	NDUFB10 Cys70&Cys77	27.47
313	VAL	ND4	2	2	ND4L Glu70	69.26	ND4 Lys237	8.09	NDUFB11 Cys112&NDUFB10 Cys148	25.28
1	MET	ND5	2	2	ND4L Glu70	111.68	ND5 His328	27.99	NDUFB10 Cys106&Cys118	18.79
159	TYR	ND5	2	2	ND4L Glu70	91.18	ND5 His248	14.50	NDUFB9 LYK (18-20)	29.58
434	GLN	ND5	2	2	ND4L Glu70	123.55	ND5 TMH12 Gly382-Tyr390	24.84	NDUFB9 LYK (18-20)	22.05
132	SER	ND6	2	2	ND3 Glu68	33.76	ND2 Glu34	37.07	NDUFS5 Cys32&Cys65	17.50
290	ARG	NDUFS2	2	2	NDUFS2 D160	22.15	ND2 Glu34	54.72	NDUFA6 LYR (34-36)	48.73
28	LEU	ND1	2	3	Q-site	9.40	ND2 Glu34	54.07	NDUFA9 NADPH	44.49
215	TYR	ND1	2	3	ND1 Glu227	11.79	ND2 Glu34	42.97	NDUFA9 NADPH	28.18
409	TYR	ND4	2	3	ND4L Glu70	79.22	ND4 TMH12 Leu367-Thr372	10.31	ND5 Cys279&NDUFB8 Cys115	35.81
236	ALA	ND5	2	3	ND4L Glu70	97.21	ND5 His248	4.69	ND5 Cys279&NDUFB8 Cys115	25.07
26	ILE	ND6	2	3	ND4L Glu34	11.07	ND2 Glu34	24.80	NDUFAB1 (α)-acyl chain	36.94
117	ASN	ND6	2	3	ND4L Glu34	32.56	ND2 Glu34	35.77	NDUFS5 Cys32&Cys65	19.50
21	TRP	NDUFB3	2	3	ND4L Glu70	133.90	ND5 TMH12 Gly382-Tyr390	39.46	NDUFB9 LYK (18-20)	24.14
179	ARG	NDUFV1	2	3	FMN	14.36	ND2 Glu34	149.61	NDUFS6 Zinc	48.60

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
366	ARG	NDUFV1	2	3	Fe-S cluster	5.61	ND2 Glu34	124.37	NDUFS6 Zinc	26.89
59	GLU	ND1	2		Q-site	15.42	ND2 Glu34	51.65	NDUFA9 NADPH	18.14
208	VAL	ND1	2		ND3-Cys39	7.33	ND2 Glu34	43.81	NDUFAB1 (α)-acyl chain	33.44
214	GLU	ND1	2		ND3-Cys39	11.75	ND2 Glu34	42.93	NDUFA9 NADPH	28.50
285	LEU	ND1	2		ND1 Glu192	12.09	ND2 Glu34	34.43	NDUFAB1 (α)-acyl chain	47.67
71	LEU	ND2	2		ND4L Glu70	10.11	ND2 Glu34	7.61	NDUFS5 Cys32&Cys65	22.31
259	GLY	ND2	2		ND4L Glu70	34.76	ND2 TMH9 Gly252-Gly259	0.00	NDUFA11 Cys94&Cys114	37.23
26	GLN	ND3	2		Q-site	20.98	ND2 Glu34	57.06	NDUFA9 NADPH	12.25
66	ASP	ND3	2		ND3 Asp66	0.00	ND2 Glu34	16.99	NDUFS5 Cys32&Cys65	35.34
32	CYS	ND4L	2		ND4L Glu34	5.53	ND2 Glu34	14.54	NDUFS5 Cys32&Cys65	34.41
145	GLU	ND5	2		ND4L Glu70	89.17	ND5 Glu145	0.00	ND5 Cys279&NDUFB8 Cys115	27.23
171	ALA	ND5	2		ND4L Glu70	91.00	ND5 His248	7.28	ND5 Cys279&NDUFB8 Cys115	23.85
237	MET	ND5	2		ND4L Glu70	100.27	ND5 His248	6.19	ND5 Cys279&NDUFB8 Cys115	26.88
250	SER	ND5	2		ND4L Glu70	105.54	ND5 His332	5.71	ND5 Cys279&NDUFB8 Cys115	23.87
465	GLY	ND5	2		ND4L Glu70	119.28	ND5 TMH12 Gly382-Tyr390	3.92	NDUFB7 Cys68&Cys79	18.83
59	TYR	ND6	2		ND6 pi-bulge (Gly61-62)	4.90	ND2 Glu34	16.01	NDUFS5 Cys32&Cys65	31.84
74	ALA	ND6	2		ND4L Glu34	16.16	ND2 Glu34	26.47	NDUFAB1 (α)-acyl chain	31.56
19	PRO	NDUFA1	2		Q-site	18.54	ND2 Glu34	47.59	NDUFA8 Cys87&Cys99	26.20
64	GLY	NDUFA10	2		ND4L Glu70	54.27	ND2 TMH9 Gly252-Gly259	36.50	NDUFA10 ADP	21.48
259	LEU	NDUFA10	2		ND4L Glu70	32.27	ND2 Leu131 pi bulge	27.22	NDUFA10 ADP	26.56
56	ARG	NDUFA13	2		ND3 Glu68	30.99	ND2 Glu34	42.80	NDUFA8 Cys87&Cys99	15.33
44	LYS	NDUFA2	2		Fe-S cluster	58.15	ND2 His112	123.85	NDUFAB1 (α)-acyl chain	58.86
108	GLU	NDUFA8	2		ND1 Glu192	52.70	ND2 Glu34	64.28	NDUFA8 Cys77&Cys109	3.80
92	GLU	NDUFB11	2		ND4L Glu70	88.26	ND4 TMH12 Leu367-Thr372	31.73	NDUFB11 Cys112&NDUFB10 Cys148	19.22
63	LEU	NDUFB9	2		ND4L Glu70	107.43	ND5 TMH7 Gly222-Pro288	37.46	NDUFAB1 (β)-acyl chain	5.28
96	TYR	NDUFB9	2		ND4L Glu70	95.21	ND5 His248	22.65	NDUFB9 LYK (18-20)	28.67
48	VAL	NDUFS1	2		Fe-S cluster	8.10	ND2 Glu34	123.68	NDUFS6 Zinc	39.22
143	GLY	NDUFS1	2		Fe-S cluster	14.43	ND2 Glu34	120.58	NDUFS6 Zinc	24.41
205	VAL	NDUFS1	2		Fe-S cluster	4.07	ND2 Glu34	111.67	NDUFS6 Zinc	22.04
499	GLN	NDUFS1	2		Fe-S cluster	37.85	ND2 TMH4 Gly107-Phe111	115.30	NDUFAB1 (α)-acyl chain	51.45
672	TYR	NDUFS1	2		Fe-S cluster	25.78	ND2 Glu34	120.72	NDUFS6 Zinc	39.03
684	MET	NDUFS1	2		Fe-S cluster	23.80	ND2 Glu34	116.89	NDUFS6 Zinc	33.81
77	ASP	NDUFS2	2		ND3-Cys39	14.15	ND2 Glu34	49.84	NDUFAB1 (α)-acyl chain	22.54
85	ARG	NDUFS2	2		NDUFS2 Y108	9.37	ND2 Glu34	68.69	NDUFAB1 (α)-acyl chain	25.92
105	ARG	NDUFS2	2		NDUFS2 Y108	5.62	ND2 Glu34	70.91	NDUFAB1 (α)-acyl chain	35.58
115	GLU	NDUFS2	2		NDUFS2 Y108	12.76	ND2 Glu34	77.67	NDUFS6 Zinc	39.06
191	ALA	NDUFS2	2		Fe-S cluster	10.58	ND2 Glu34	77.44	NDUFS6 Zinc	34.79
300	ARG	NDUFS2	2		NDUFS2 H59	16.31	ND2 Glu34	61.57	NDUFAB1 (α)-acyl chain	48.26

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
410	MET	NDUFS2	2		NDUFS2 D160	14.09	ND2 Glu34	39.72	NDUFAB1 (α)-acyl chain	35.71
413	ASP	NDUFS2	2		NDUFS2 D160	10.37	ND2 Glu34	43.80	NDUFAB1 (α)-acyl chain	37.82
187	PRO	NDUFS3	2		Fe-S cluster	24.02	ND2 Glu34	74.42	NDUFAB1 (α)-acyl chain	18.57
72	TRP	NDUFS4	2		Fe-S cluster	16.47	ND2 Glu34	87.51	NDUFS6 Zinc	33.45
77	ASP	NDUFS4	2		Fe-S cluster	25.77	ND2 Glu34	82.61	NDUFAB1 (α)-acyl chain	23.44
95	PRO	NDUFS5	2		ND4L Glu34	51.54	ND2 Glu34	55.11	NDUFA8 Cys45&Cys55	29.91
67	ARG	NDUFS7	2		Q-site	13.20	ND2 Glu34	62.05	NDUFA9 NADPH	35.78
107	ARG	NDUFS7	2		Q-site	16.68	ND2 Glu34	58.60	NDUFA9 NADPH	18.11
20	ARG	NDUFS8	2		ND1 Glu192	28.74	ND2 Glu34	48.62	NDUFA8 Cys87&Cys99	53.19
60	ARG	NDUFS8	2		Fe-S cluster	13.53	ND2 Glu34	85.32	NDUFS6 Zinc	27.42
120	GLY	NDUFS8	2		Fe-S cluster	3.81	ND2 Glu34	84.17	NDUFS6 Zinc	26.26
36	SER	NDUFV1	2		FMN	24.18	ND2 Glu34	169.01	NDUFS6 Zinc	70.94
68	ARG	NDUFV1	2		FMN	3.70	ND2 Glu34	146.18	NDUFS6 Zinc	54.97
91	LYS	NDUFV1	2		FMN	16.92	ND2 Glu34	148.30	NDUFS6 Zinc	59.29
127	ARG	NDUFV1	2		FMN	16.51	ND2 Glu34	155.57	NDUFS6 Zinc	66.14
184	TYR	NDUFV1	2		FMN	5.48	ND2 Glu34	139.57	NDUFS6 Zinc	39.78
191	ALA	NDUFV1	2		FMN	8.30	ND2 Glu34	137.31	NDUFS6 Zinc	42.02
226	GLU	NDUFV1	2		FMN	7.24	ND2 Glu34	151.35	NDUFS6 Zinc	57.63
232	PRO	NDUFV1	2		FMN	15.21	ND2 Glu34	159.85	NDUFS6 Zinc	67.89
321	ALA	NDUFV1	2		FMN	17.92	ND2 Glu34	141.01	NDUFS6 Zinc	57.69
357	GLU	NDUFV1	2		Fe-S cluster	7.92	ND2 Glu34	134.23	NDUFS6 Zinc	32.84
358	SER	NDUFV1	2		Fe-S cluster	4.37	ND2 Glu34	131.67	NDUFS6 Zinc	32.53
412	ALA	NDUFV1	2		Fe-S cluster	10.93	ND2 Glu34	127.62	NDUFS6 Zinc	36.25
177	LYS	NDUFV2	2		Fe-S cluster	24.84	ND2 Glu34	160.54	NDUFS6 Zinc	51.37
4	ALA	ND1	3	1	Q-site	22.57	ND2 Glu34	49.33	NDUFA8 Cys87&Cys99	28.40
304	TYR	ND1	3	1	ND3 Glu68	16.56	ND2 Glu34	31.99	NDUFA8 Cys87&Cys99	33.11
150	ASN	ND2	3	1	ND4L Glu70	25.97	ND2 TMH4 Gly107-Phe111	14.20	NDUFS5 Cys42&Cys55	23.64
331	ALA	ND2	3	1	ND4L Glu70	39.18	ND2 TMH9 Gly252-Gly259	7.50	NDUFB10 Cys70&Cys77	46.11
458	ALA	ND5	3	1	ND4L Glu70	119.87	ND5 TMH12 Gly382-Tyr390	9.18	NDUFB7 Cys68&Cys79	27.84
544	THR	ND5	3	1	ND4L Glu70	73.97	ND5 Asp554 pi bulge	11.18	NDUFA11 Cys94&Cys114	28.58
119	ASN	ND6	3	1	ND4L Glu34	35.24	ND2 Glu34	39.23	NDUFS5 Cys32&Cys65	22.01
11	VAL	ND1	3	2	Q-site	11.63	ND2 Glu34	48.05	NDUFA8 Cys87&Cys99	34.57
39	VAL	ND1	3	2	Q-site	16.08	ND2 Glu34	64.86	NDUFA9 NADPH	34.68
64	ALA	ND1	3	2	ND3-Cys39	17.40	ND2 Glu34	43.73	NDUFA9 NADPH	24.62
113	VAL	ND1	3	2	ND1 Glu143	6.45	ND2 Glu34	27.93	NDUFS5 Cys32&Cys65	43.39
144	VAL	ND1	3	2	ND1 Glu143	3.80	ND2 Glu34	23.51	NDUFS5 Cys32&Cys65	40.50
147	ALA	ND1	3	2	ND1 Glu143	6.07	ND2 Glu34	25.01	NDUFA8 Cys87&Cys99	36.99
187	ILE	ND1	3	2	ND1 Glu192	8.79	ND2 Glu34	36.69	NDUFA8 Cys87&Cys99	40.94
64	ALA	ND2	3	2	ND4L Glu70	4.74	ND2 Glu34	6.39	NDUFS5 Cys32&Cys65	33.21

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
122	THR	ND2	3	2	ND4L Glu70	21.41	ND2 Leu131 pi bulge	11.09	NDUFA10 ADP	35.82
10	ASN	ND3	3	2	ND1 Glu227	16.59	ND2 Glu34	39.26	NDUFA9 NADPH	38.63
423	ILE	ND4	3	2	ND4L Glu70	80.22	ND4 Ala223 pi bulge	24.72	NDUFB9 LYK (18-20)	36.26
9	THR	ND5	3	2	ND4L Glu70	104.03	ND4 TMH12 Leu367-Thr372	19.14	NDUFB10 Cys106&Cys118	23.76
21	THR	ND5	3	2	ND4L Glu70	98.62	ND4 TMH12 Leu367-Thr372	20.13	NDUFB10 Cys106&Cys118	41.82
100	ILE	ND5	3	2	ND4L Glu70	108.53	ND5 Lys336	12.70	ND5 Cys279&NDUFB8 Cys115	34.64
267	ALA	ND5	3	2	ND4L Glu70	100.78	ND5 TMH7 Gly222-Pro258	17.43	ND5 Cys279&NDUFB8 Cys115	15.93
398	THR	ND5	3	2	ND4L Glu70	118.28	ND5 TMH12 Gly382-Tyr390	9.96	ND5 Cys279&NDUFB8 Cys115	17.46
576	ILE	ND5	3	2	ND4L Glu70	33.21	ND2 His112	10.93	NDUFA11 Cys94&Cys114	22.86
33	ILE	ND6	3	2	ND4L Glu34	9.78	ND2 Glu34	21.97	NDUFS5 Cys32&Cys65	34.81
58	ILE	ND6	3	2	ND6 pi-bulge (Gly61-62)	4.93	ND2 Glu34	16.18	NDUFS5 Cys32&Cys65	31.61
171	ALA	ND6	3	2	ND4L Glu70	14.83	ND2 Glu34	20.12	NDUFA10 ADP	42.98
32	GLY	NDUFA1	3	2	Q-site	38.15	ND2 Glu34	54.76	NDUFA8 Cys87&Cys99	13.43
103	ARG	NDUFA11	3	2	ND4L Glu70	50.28	ND2 His112	29.59	NDUFA11 Cys94&Cys114	14.12
40	ARG	NDUFA9	3	2	Fe-S cluster	27.78	ND2 Glu34	74.28	NDUFA9 NADPH	17.27
408	LEU	NDUFS1	3	2	Fe-S cluster	32.88	ND2 TMH4 Gly107-Phe111	136.61	NDUFS6 Zinc	55.40
2	PRO	ND1	3	3	Q-site	25.33	ND2 Glu34	45.79	NDUFA8 Cys87&Cys99	27.17
56	PHE	ND1	3	3	Q-site	9.72	ND2 Glu34	50.90	NDUFA9 NADPH	23.11
128	ALA	ND1	3	3	ND3-Cys39	11.28	ND2 Glu34	36.90	NDUFAB1 (α)-acyl chain	31.52
176	LEU	ND1	3	3	ND1 Glu192	22.24	ND2 Glu34	43.61	NDUFA8 Cys87&Cys99	28.96
230	ASN	ND1	3	3	ND1 Glu227	5.41	ND2 Glu34	37.50	NDUFA8 Cys87&Cys99	37.80
276	ALA	ND1	3	3	Q-site	12.68	ND2 Glu34	49.92	NDUFA8 Cys87&Cys99	50.52
328	THR	ND2	3	3	ND4L Glu70	36.70	ND2 TMH9 Gly252-Gly259	11.53	NDUFA10 ADP	46.02
71	ALA	ND4L	3	3	ND4L Glu70	3.79	ND2 Glu34	9.89	NDUFS5 Cys32&Cys65	35.68
99	SER	ND5	3	3	ND4L Glu70	111.75	ND5 Lys336	12.69	NDUFB7 Cys68&Cys79	32.95
149	ILE	ND5	3	3	ND4L Glu70	89.06	ND5 Glu145	6.11	ND5 Cys279&NDUFB8 Cys115	30.12
112	VAL	ND6	3	3	ND4L Glu34	33.15	ND2 TMH4 Gly107-Phe111	28.98	NDUFS5 Cys32&Cys65	18.59
100	LEU	NDUFA10	3	3	ND4L Glu70	51.24	ND2 Leu131 pi bulge	38.28	NDUFA10 ADP	6.64
251	GLN	NDUFA10	3	3	ND3 Glu68	30.62	ND2 Leu131 pi bulge	33.50	NDUFA10 ADP	31.23
50	ARG	NDUFA9	3	3	Q-site	33.71	ND2 Glu34	74.53	NDUFA9 NADPH	5.81
279	ILE	NDUFA9	3	3	Q-site	36.45	ND2 TMH4 Gly107-Phe111	59.08	NDUFA9 NADPH	15.06
157	ARG	NDUFB9	3	3	ND4L Glu70	102.56	ND4 Ala223 pi bulge	54.90	NDUFAB1 (β)-acyl chain	31.94
154	ILE	NDUFS1	3	3	Fe-S cluster	4.69	ND2 Glu34	117.52	NDUFS6 Zinc	22.67
181	MET	NDUFS1	3	3	Fe-S cluster	10.05	ND2 Glu34	126.93	NDUFS6 Zinc	25.62
483	VAL	NDUFS1	3	3	Fe-S cluster	43.34	ND2 His112	133.15	NDUFAB1 (α)-acyl chain	66.41
511	VAL	NDUFS1	3	3	Fe-S cluster	48.75	ND2 TMH4 Gly107-Phe111	121.34	NDUFAB1 (α)-acyl chain	60.09
287	VAL	NDUFS2	3	3	NDUFS2 H59	23.71	ND2 Glu34	63.18	NDUFAB1 (α)-acyl chain	51.08

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
347	HIS	NDUFS2	3	3	Fe-S cluster	10.24	ND2 Glu34	99.43	NDUFS6 Zinc	22.25
210	ARG	NDUFS3	3	3	Fe-S cluster	21.75	ND2 Glu34	98.98	NDUFS6 Zinc	44.15
58	ALA	NDUFS6	3	3	Fe-S cluster	16.49	ND2 Glu34	114.68	NDUFS6 Zinc	7.85
60	ASP	NDUFS6	3	3	Fe-S cluster	14.69	ND2 Glu34	108.34	NDUFS6 Zinc	5.15
167	ARG	NDUFS7	3	3	Q-site	24.00	ND2 Glu34	68.70	NDUFA9 NADPH	12.53
133	ALA	NDUFV1	3	3	FMN	17.71	ND2 Glu34	154.23	NDUFS6 Zinc	60.73
337	MET	NDUFV1	3	3	Fe-S cluster	12.43	ND2 Glu34	142.09	NDUFS6 Zinc	47.31
12	PRO	ND1	3		Q-site	12.41	ND2 Glu34	48.79	NDUFA8 Cys87&Cys99	32.58
25	ARG	ND1	3		Q-site	7.57	ND2 Glu34	53.83	NDUFA9 NADPH	40.15
34	ARG	ND1	3		Q-site	14.52	ND2 Glu34	62.00	NDUFA9 NADPH	44.39
38	ASN	ND1	3		Q-site	15.48	ND2 Glu34	63.30	NDUFA9 NADPH	37.82
218	GLY	ND1	3		Q-site	10.58	ND2 Glu34	46.45	NDUFA9 NADPH	25.17
279	ARG	ND1	3		NDUFS2 D160	9.72	ND2 Glu34	41.89	NDUFAB1 (α)-acyl chain	45.73
55	ALA	ND2	3		ND4L Glu70	16.32	ND2 Leu131 pi bulge	13.07	NDUFA10 ADP	41.34
60	PHE	ND2	3		ND4L Glu70	8.41	ND2 Lys105	9.21	NDUFS5 Cys32&Cys65	38.86
128	LEU	ND2	3		ND4L Glu70	21.87	ND2 Leu131 pi bulge	5.09	NDUFA10 ADP	41.42
158	LEU	ND4	3		ND4L Glu70	43.19	ND4 Glu123	6.25	NDUFA11 Cys94&Cys114	25.09
206	LYS	ND4	3		ND4L Glu70	55.39	ND4 Lys206	0.00	NDUFA11 Cys94&Cys114	28.06
372	THR	ND4	3		ND4L Glu70	81.05	ND4 TMH12 Leu367-Thr372	0.00	NDUFB10 Cys106&Cys118	32.25
387	SER	ND4	3		ND4L Glu70	78.21	ND4 Glu378	14.24	NDUFB11 Cys112&NDUFB10 Cys148	22.15
412	THR	ND4	3		ND4L Glu70	82.83	ND4 TMH12 Leu367-Thr372	12.42	NDUFB9 LYK (18-20)	34.05
25	HIS	ND4L	3		ND4L Glu34	14.16	ND2 TMH4 Gly107-Phe111	19.56	NDUFA6 LYR (34-36)	36.91
41	PHE	ND4L	3		ND4L Glu34	10.51	ND2 Glu34	12.41	NDUFS5 Cys32&Cys65	21.81
96	VAL	ND5	3		ND4L Glu70	109.85	ND5 Lys336	10.51	NDUFB7 Cys68&Cys79	30.15
239	GLY	ND5	3		ND4L Glu70	98.96	ND5 His248	7.57	NDUFB9 LYK (18-20)	28.89
243	VAL	ND5	3		ND4L Glu70	101.56	ND5 His248	6.09	ND5 Cys279&NDUFB8 Cys115	30.69
253	VAL	ND5	3		ND4L Glu70	100.10	ND5 TMH7 Gly222-Pro242	4.62	ND5 Cys279&NDUFB8 Cys115	19.22
312	LEU	ND5	3		ND4L Glu70	108.91	ND5 His328	6.17	ND5 Cys279&NDUFB8 Cys115	11.14
348	HIS	ND5	3		ND4L Glu70	110.87	ND5 His248	18.70	NDUFB9 LYK (18-20)	24.81
392	LYS	ND5	3		ND4L Glu70	117.19	ND5 Lys392	0.00	ND5 Cys279&NDUFB8 Cys115	19.17
415	ALA	ND5	3		ND4L Glu70	113.63	ND5 Lys392	8.20	ND5 Cys279&NDUFB8 Cys115	14.50
499	LEU	ND5	3		ND4L Glu70	125.67	ND5 TMH12 Gly382-Tyr390	14.66	ND5 Cys279&NDUFB8 Cys115	28.37
505	ASN	ND5	3		ND4L Glu70	120.74	ND5 TMH12 Gly382-Tyr390	21.49	NDUFAB1 (β)-acyl chain	18.17
585	LYS	ND5	3		ND4L Glu34	19.34	ND2 TMH4 Gly107-Phe111	12.58	NDUFA11 Cys94&Cys114	38.52
25	PRO	ND6	3		ND4L Glu34	10.09	ND2 Glu34	22.70	NDUFAB1 (α)-acyl chain	37.33
79	PRO	ND6	3		ND4L Glu34	19.05	ND2 TMH4 Gly107-Phe111	30.39	NDUFAB1 (α)-acyl chain	28.90

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
82	TRP	ND6	3		ND4L Glu34	20.01	ND2 TMH4 Gly107-Phe111	29.73	NDUFAB1 (α)-acyl chain	34.90
101	GLY	ND6	3		ND4L Glu34	25.56	ND2 TMH4 Gly107-Phe111	25.79	NDUFS5 Cys32&Cys65	29.99
43	TYR	NDUFA1	3		ND3 Glu68	33.17	ND2 Glu34	38.51	NDUFA8 Cys87&Cys99	20.64
173	ASP	NDUFA10	3		NDUFS2 H59	52.46	ND2 Leu131 pi bulge	56.39	NDUFA10 ADP	13.67
176	VAL	NDUFA10	3		NDUFS2 H59	60.40	ND2 Leu131 pi bulge	58.73	NDUFA10 ADP	13.95
9	TRP	NDUFA11	3		ND4L Glu70	47.28	ND2 TMH4 Gly107-Phe111	28.47	NDUFA11 Cys17-Cy74	12.17
77	ALA	NDUFA11	3		ND4L Glu70	58.84	ND4 TMH7 Met207-Leu214	31.94	NDUFA11 Cys17-Cy74	5.23
81	GLU	NDUFA11	3		ND4L Glu70	63.62	ND4 TMH7 Met207-Leu214	33.69	NDUFA11 Cys17-Cy74	10.48
68	TYR	NDUFA2	3		Fe-S cluster	61.42	ND2 His112	107.55	NDUFAB1 (α)-acyl chain	45.81
26	ALA	NDUFA9	3		ND3-Cys39	31.70	ND2 Glu34	72.16	NDUFA9 NADPH	5.47
33	TYR	NDUFA9	3		ND3-Cys39	28.38	ND2 TMH4 Gly107-Phe111	68.88	NDUFA9 NADPH	8.10
287	MET	NDUFA9	3		ND3-Cys39	42.83	ND2 TMH4 Gly107-Phe111	62.64	NDUFA9 NADPH	10.39
303	LEU	NDUFA9	3		ND3-Cys39	44.53	ND2 TMH4 Gly107-Phe111	78.40	NDUFA9 NADPH	17.77
60	THR	NDUFB9	3		ND4L Glu70	106.36	ND5 TMH7 Gly222-Pro288	41.00	NDUFAB1 (β)-acyl chain	7.62
39	ARG	NDUFS1	3		Fe-S cluster	3.72	ND2 Glu34	113.46	NDUFS6 Zinc	32.84
230	VAL	NDUFS1	3		Fe-S cluster	28.23	ND2 TMH4 Gly107-Phe111	110.75	NDUFS6 Zinc	44.06
286	ASN	NDUFS1	3		Fe-S cluster	49.51	ND2 TMH4 Gly107-Phe111	104.99	NDUFA9 NADPH	37.38
432	ILE	NDUFS1	3		Fe-S cluster	44.92	ND2 His112	143.29	NDUFS6 Zinc	68.45
452	VAL	NDUFS1	3		Fe-S cluster	34.95	ND2 His112	129.63	NDUFS6 Zinc	58.10
543	ILE	NDUFS1	3		Fe-S cluster	51.66	ND2 TMH4 Gly107-Phe111	110.68	NDUFA9 NADPH	48.03
553	GLY	NDUFS1	3		Fe-S cluster	43.60	ND2 TMH4 Gly107-Phe111	112.85	NDUFA9 NADPH	50.23
577	GLU	NDUFS1	3		Fe-S cluster	32.27	ND2 His112	117.01	NDUFAB1 (α)-acyl chain	51.57
680	ALA	NDUFS1	3		Fe-S cluster	30.93	ND2 Glu34	113.18	NDUFS6 Zinc	38.46
20	TYR	NDUFS2	3		ND4L Glu70	39.62	ND2 His112	18.85	NDUFA11 Cys94&Cys114	27.98
376	VAL	NDUFS2	3		NDUFS2 Y108	17.11	ND2 Glu34	76.01	NDUFAB1 (α)-acyl chain	37.56
42	ARG	NDUFS4	3		Fe-S cluster	18.01	ND2 Glu34	104.58	NDUFS6 Zinc	35.96
49	VAL	NDUFS6	3		Fe-S cluster	23.83	ND2 Glu34	118.27	NDUFS6 Zinc	11.87
55	ARG	NDUFS6	3		Fe-S cluster	21.62	ND2 Glu34	124.51	NDUFS6 Zinc	15.25
74	ASN	NDUFS6	3		Fe-S cluster	16.26	ND2 Glu34	120.76	NDUFS6 Zinc	12.63
77	ARG	NDUFS7	3		ND3-Cys39	11.19	ND2 Glu34	48.95	NDUFA9 NADPH	25.02
81	VAL	NDUFS7	3		Q-site	14.31	ND2 Glu34	59.09	NDUFA9 NADPH	23.48
71	PRO	NDUFS8	3		Fe-S cluster	15.18	ND2 Glu34	102.71	NDUFS6 Zinc	12.47
81	LYS	NDUFS8	3		Fe-S cluster	4.81	ND2 Glu34	95.31	NDUFS6 Zinc	19.95
109	TYR	NDUFS8	3		Fe-S cluster	8.76	ND2 Glu34	91.70	NDUFS6 Zinc	22.93
127	PRO	NDUFS8	3		Fe-S cluster	5.82	ND2 Glu34	98.04	NDUFS6 Zinc	13.76
170	GLN	NDUFS8	3		Fe-S cluster	18.02	ND2 Glu34	95.55	NDUFS6 Zinc	20.11
19	ASP	NDUFV1	3		FMN	21.99	ND2 Glu34	162.88	NDUFS6 Zinc	71.24
75	THR	NDUFV1	3		FMN	6.01	ND2 Glu34	146.28	NDUFS6 Zinc	59.43
131	ALA	NDUFV1	3		FMN	16.48	ND2 Glu34	151.99	NDUFS6 Zinc	62.11
140	GLY	NDUFV1	3		FMN	12.36	ND2 Glu34	148.02	NDUFS6 Zinc	45.30

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
247	ARG	NDUFV1	3		FMN	18.21	ND2 Glu34	149.15	NDUFS6 Zinc	65.90
263	ASN	NDUFV1	3		Fe-S cluster	17.14	ND2 Glu34	154.36	NDUFS6 Zinc	57.69
297	VAL	NDUFV1	3		Fe-S cluster	14.78	ND2 Glu34	140.75	NDUFS6 Zinc	48.57
348	ALA	NDUFV1	3		Fe-S cluster	13.37	ND2 Glu34	131.13	NDUFS6 Zinc	33.24
153	MET	NDUFV2	3		Fe-S cluster	5.01	ND2 Glu34	150.27	NDUFS6 Zinc	46.19
158	ASP	NDUFV2	3		Fe-S cluster	14.90	ND2 Glu34	154.70	NDUFS6 Zinc	48.81
67	THR	ND1		1	ND1 Glu227	20.84	ND2 Glu34	36.53	NDUFAB1 (α)-acyl chain	30.71
81	ILE	ND1		1	ND1 Glu227	11.26	ND2 Glu34	32.67	NDUFA8 Cys87&Cys99	38.80
309	ILE	ND1		1	ND3 Glu68	21.02	ND2 Glu34	34.38	NDUFA8 Cys87&Cys99	30.88
8	VAL	ND2		1	ND3 Glu68	15.88	ND2 Glu34	15.63	NDUFS5 Cys32&Cys65	42.70
119	THR	ND2		1	ND4L Glu70	20.63	ND2 Leu131 pi bulge	6.76	NDUFA10 ADP	39.68
193	VAL	ND2		1	ND4L Glu70	28.34	ND2 TMH4 Gly107-Phe111	10.63	NDUFS5 Cys42&Cys55	29.49
237	LEU	ND2		1	ND4L Glu70	44.62	ND2 TMH9 Gly252-Gly259	20.42	NDUFA10 ADP	29.38
325	PHE	ND2		1	ND4L Glu70	33.67	ND2 TMH9 Gly252-Gly259	15.55	NDUFA10 ADP	42.73
88	VAL	ND3		1	ND3 Glu68	14.18	ND2 Glu34	26.91	NDUFS5 Cys32&Cys65	33.39
114	THR	ND3		1	ND3 Asp66	21.37	ND2 Glu34	28.72	NDUFA6 LYR (34-36)	42.15
140	PRO	ND4		1	ND4L Glu70	55.01	ND4 Ala223 pi bulge	11.72	NDUFA11 Cys94&Cys114	36.53
404	ALA	ND4		1	ND4L Glu70	79.98	ND4 TMH12 Leu367-Thr372	4.17	ND5 Cys279&NDUFB8 Cys115	30.44
47	MET	ND4L		1	ND4L Glu34	20.33	ND2 Glu34	19.29	NDUFS5 Cys32&Cys65	11.90
8	THR	ND5		1	ND4L Glu70	107.32	ND4 TMH12 Leu367-Thr372	22.07	NDUFB10 Cys106&Cys118	23.66
24	VAL	ND5		1	ND4L Glu70	96.39	ND4 TMH12 Leu367-Thr372	23.29	NDUFB10 Cys106&Cys118	46.35
257	ILE	ND5		1	ND4L Glu70	101.38	ND5 TMH7 Gly222-Pro246	5.59	ND5 Cys279&NDUFB8 Cys115	15.09
314	MET	ND5		1	ND4L Glu70	104.67	ND5 TMH7 Gly222-Pro258	5.92	ND5 Cys279&NDUFB8 Cys115	9.25
475	ALA	ND5		1	ND4L Glu70	116.65	ND5 TMH12 Gly382-Tyr390	13.80	NDUFB7 Cys58-89	14.41
482	ILE	ND5		1	ND4L Glu70	126.16	ND5 TMH12 Gly382-Tyr390	12.56	NDUFB7 Cys58-89	16.77
485	TYR	ND5		1	ND4L Glu70	127.33	ND5 TMH12 Gly382-Tyr390	12.86	ND5 Cys279&NDUFB8 Cys115	23.04
515	SER	ND5		1	ND4L Glu70	116.07	ND5 His230	24.88	NDUFAB1 (β)-acyl chain	12.12
531	SER	ND5		1	ND4L Glu70	93.56	ND5 TMH7 Gly222-Pro289	9.65	NDUFAB1 (β)-acyl chain	22.40
533	THR	ND5		1	ND4L Glu70	90.02	ND5 TMH7 Gly222-Pro289	7.67	ND5 Cys279&NDUFB8 Cys115	21.07
555	LEU	ND5		1	ND4L Glu70	59.50	ND5 Asp554 pi bulge	3.80	NDUFA11 Cys94&Cys114	22.57
166	ILE	ND6		1	ND4L Glu70	10.73	ND2 Glu34	13.36	NDUFS5 Cys32&Cys65	42.42
47	LYS	NDUFA7		1	Fe-S cluster	24.36	ND2 Glu34	107.83	NDUFS6 Zinc	30.26
259	PRO	NDUFA9		1	ND3-Cys39	51.07	ND2 TMH4 Gly107-Phe111	58.83	NDUFA9 NADPH	23.50
145	PRO	NDUFB9		1	ND4L Glu70	87.77	ND5 Asp554 pi bulge	46.64	NDUFAB1 (β)-acyl chain	33.51
46	LEU	NDUFC2		1	ND4L Glu70	45.34	ND2 TMH9 Gly252-Gly259	17.40	NDUFA10 ADP	37.87

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
319	PRO	NDUFS2		1	NDUFS2 Y108	30.93	ND2 Glu34	91.15	NDUFS6 Zinc	40.20
102	VAL	NDUFV2		1	Fe-S cluster	7.97	ND2 Glu34	146.25	NDUFS6 Zinc	40.47
10	ILE	ND1		2	Q-site	13.63	ND2 Glu34	44.61	NDUFA8 Cys87&Cys99	32.88
29	GLY	ND1		2	Q-site	11.15	ND2 Glu34	57.66	NDUFA9 NADPH	43.66
43	TYR	ND1		2	Q-site	16.17	ND2 Glu34	62.98	NDUFA9 NADPH	39.79
69	THR	ND1		2	ND1 Glu227	19.68	ND2 Glu34	35.61	NDUFA9 NADPH	32.36
73	THR	ND1		2	ND1 Glu227	15.45	ND2 Glu34	35.08	NDUFA9 NADPH	34.22
87	THR	ND1		2	ND1 Glu192	15.86	ND2 Glu34	36.85	NDUFA8 Cys87&Cys99	28.71
89	LEU	ND1		2	ND1 Glu192	14.78	ND2 Glu34	38.05	NDUFA8 Cys87&Cys99	25.50
229	THR	ND1		2	ND1 Glu227	5.86	ND2 Glu34	40.53	NDUFA8 Cys87&Cys99	37.50
236	THR	ND1		2	ND1 Glu192	11.40	ND2 Glu34	40.60	NDUFA8 Cys87&Cys99	29.33
248	ASP	ND1		2	ND1 Glu192	30.07	ND2 Glu34	51.86	NDUFA8 Cys87&Cys99	11.85
263	THR	ND1		2	ND1 Glu192	15.16	ND2 Glu34	45.87	NDUFA8 Cys87&Cys99	30.54
275	THR	ND1		2	Q-site	10.44	ND2 Glu34	50.03	NDUFA9 NADPH	48.29
307	MET	ND1		2	ND3 Glu68	20.10	ND2 Glu34	35.20	NDUFA8 Cys87&Cys99	32.89
13	ILE	ND2		2	ND4L Glu70	14.10	ND2 Glu34	10.00	NDUFS5 Cys32&Cys65	38.04
31	VAL	ND2		2	ND4L Glu70	8.66	ND2 Glu34	4.88	NDUFS5 Cys32&Cys65	26.60
43	VAL	ND2		2	ND4L Glu70	15.72	ND2 Leu131 pi bulge	13.25	NDUFA10 ADP	41.45
76	PHE	ND2		2	ND4L Glu70	18.50	ND2 Glu34	16.12	NDUFS5 Cys32&Cys65	16.03
80	LEU	ND2		2	ND4L Glu70	23.70	ND2 Glu34	21.18	NDUFS5 Cys32&Cys65	11.55
88	ASN	ND2		2	ND4L Glu70	24.52	ND2 Lys105	19.01	NDUFS5 Cys42&Cys55	14.70
89	THR	ND2		2	ND4L Glu70	25.12	ND2 Lys105	20.45	NDUFS5 Cys42&Cys55	12.99
100	MET	ND2		2	ND4L Glu70	16.00	ND2 TMH4 Gly107-Phe111	5.37	NDUFS5 Cys32&Cys65	30.13
152	SER	ND2		2	ND4L Glu70	28.63	ND2 TMH4 Gly107-Phe111	11.93	NDUFS5 Cys42&Cys55	28.53
164	ALA	ND2		2	ND4L Glu70	27.14	ND2 His112	4.76	NDUFA11 Cys94&Cys114	25.87
202	ILE	ND2		2	ND4L Glu70	26.46	ND2 Lys135	11.01	NDUFS5 Cys42&Cys55	31.31
203	LEU	ND2		2	ND4L Glu70	29.42	ND2 Lys263	10.19	NDUFS5 Cys42&Cys55	34.21
265	ALA	ND2		2	ND4L Glu70	36.45	ND2 Lys263	6.01	NDUFB10 Cys70&Cys77	33.53
270	PHE	ND2		2	ND4L Glu70	35.95	ND2 Lys263	10.98	NDUFA11 Cys94&Cys114	31.39
278	ILE	ND2		2	ND4L Glu70	35.54	ND2 Lys263	13.04	NDUFA11 Cys94&Cys114	23.70
284	THR	ND2		2	ND4L Glu70	38.44	ND2 Lys263	10.75	NDUFA11 Cys94&Cys114	20.92
288	LEU	ND2		2	ND4L Glu70	36.61	ND2 TMH9 Gly252-Gly259	9.80	NDUFA11 Cys94&Cys114	23.62
323	THR	ND2		2	ND4L Glu70	33.67	ND2 Leu131 pi bulge	18.82	NDUFA10 ADP	38.84
342	PHE	ND2		2	ND4L Glu70	39.16	ND2 TMH9 Gly252-Gly259	12.06	NDUFB10 Cys70&Cys77	34.30
9	ILE	ND3		2	Q-site	19.16	ND2 Glu34	41.22	NDUFA8 Cys87&Cys99	37.62
29	GLY	ND3		2	ND3-Cys39	19.72	ND2 Glu34	54.22	NDUFA9 NADPH	14.52
49	VAL	ND3		2	ND3-Cys39	18.54	ND2 Glu34	33.50	NDUFAB1 (α)-acyl chain	25.57
96	ILE	ND3		2	ND3 Glu68	6.97	ND2 Glu34	22.73	NDUFS5 Cys32&Cys65	39.75
103	ALA	ND3		2	ND3 Glu68	11.93	ND2 Glu34	24.73	NDUFS5 Cys32&Cys65	47.49
47	ASN	ND4		2	ND4L Glu70	69.72	ND4 Lys237	20.73	NDUFB11 Cys112&NDUFB10 Cys148	16.34

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
50	PHE	ND4		2	ND4L Glu70	66.68	ND4 Lys237	23.18	NDUFB10 Cys70&Cys77	15.76
54	PRO	ND4		2	ND4L Glu70	57.45	ND2 TMH9 Gly252-Gly259	21.17	NDUFB10 Cys70&Cys77	18.77
86	SER	ND4		2	ND4L Glu70	65.33	ND4 Ala223 pi bulge	18.42	NDUFA10 ADP	37.87
89	LEU	ND4		2	ND4L Glu70	59.43	ND4 Ala223 pi bulge	20.84	NDUFA10 ADP	38.11
101	SER	ND4		2	ND4L Glu70	53.84	ND2 TMH9 Gly252-Gly259	10.42	NDUFB10 Cys70&Cys77	40.82
110	PHE	ND4		2	ND4L Glu70	58.85	ND4 Lys237	11.13	NDUFB10 Cys70&Cys77	28.07
131	ALA	ND4		2	ND4L Glu70	50.12	ND4 Ala223 pi bulge	10.50	NDUFA11 Cys94&Cys114	38.81
138	ASN	ND4		2	ND4L Glu70	58.86	ND4 Ala223 pi bulge	7.37	NDUFA11 Cys94&Cys114	40.73
149	PHE	ND4		2	ND4L Glu70	50.24	ND4 Lys206	7.65	NDUFA11 Cys94&Cys114	27.77
230	VAL	ND4		2	ND4L Glu70	61.28	ND4 His220	7.24	NDUFA11 Cys94&Cys114	40.39
299	THR	ND4		2	ND4L Glu70	67.26	ND4 TMH7 Met207-Leu214	7.93	NDUFB11 Cys112&NDUFB10 Cys148	25.99
390	ASN	ND4		2	ND4L Glu70	79.13	ND4 Glu378	16.48	ND5 Cys279&NDUFB8 Cys115	27.22
402	VAL	ND4		2	ND4L Glu70	77.36	ND5 Arg176	7.24	ND5 Cys279&NDUFB8 Cys115	29.93
424	ASN	ND4		2	ND4L Glu70	80.92	ND4 Ala223 pi bulge	25.66	NDUFB9 LYK (18-20)	38.65
442	SER	ND4		2	ND4L Glu70	79.33	ND4 TMH12 Leu367-Thr372	10.39	NDUFB10 Cys106&Cys118	40.66
445	LEU	ND4		2	ND4L Glu70	81.14	ND4 TMH12 Leu367-Thr372	6.42	NDUFB10 Cys106&Cys118	35.62
459	SER	ND4		2	ND4L Glu70	76.59	ND4 Glu378	17.16	NDUFB11 Cys112&NDUFB10 Cys148	16.18
13	THR	ND4L		2	ND4L Glu34	13.15	ND2 TMH4 Gly107-Phe111	14.55	NDUFS5 Cys32&Cys65	27.90
62	ALA	ND4L		2	ND3 Glu68	11.14	ND2 Glu34	8.73	NDUFS5 Cys32&Cys65	23.70
94	ASN	ND4L		2	ND4L Glu70	24.26	ND2 His112	16.39	NDUFA6 LYR (34-36)	34.25
4	HIS	ND5		2	ND4L Glu70	111.18	ND5 His328	25.63	NDUFB10 Cys106&Cys118	21.65
13	THR	ND5		2	ND4L Glu70	101.42	ND4 TMH12 Leu367-Thr372	16.65	NDUFB10 Cys106&Cys118	28.77
22	THR	ND5		2	ND4L Glu70	95.91	ND4 TMH12 Leu367-Thr372	19.26	NDUFB10 Cys106&Cys118	43.55
23	LEU	ND5		2	ND4L Glu70	94.19	ND4 TMH12 Leu367-Thr372	19.51	NDUFB10 Cys106&Cys118	43.48
52	THR	ND5		2	ND4L Glu70	114.54	ND5 His328	17.59	NDUFB7 Cys68&Cys79	15.97
69	ALA	ND5		2	ND4L Glu70	82.80	ND4 TMH12 Leu367-Thr372	11.78	NDUFB10 Cys106&Cys118	23.52
74	THR	ND5		2	ND4L Glu70	85.31	ND4 TMH12 Leu367-Thr372	16.32	NDUFB10 Cys106&Cys118	17.21
182	PHE	ND5		2	ND4L Glu70	88.63	ND5 Asp179	4.50	ND5 Cys279&NDUFB8 Cys115	20.47
190	ILE	ND5		2	ND4L Glu70	85.68	ND4 Glu378	15.06	NDUFB10 Cys106&Cys118	18.33
201	MET	ND5		2	ND4L Glu70	94.48	ND5 Asp179	19.75	NDUFB10 Cys106&Cys118	11.09
202	ALA	ND5		2	ND4L Glu70	98.18	ND5 Asp179	21.33	NDUFB10 Cys106&Cys118	10.07

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
205	ASN	ND5		2	ND4L Glu70	94.77	ND5 Asp179	21.70	NDUFB10 Cys106&Cys118	12.72
261	ILE	ND5		2	ND4L Glu70	103.19	ND5 His328	10.72	ND5 Cys279&NDUFB8 Cys115	18.09
265	PRO	ND5		2	ND4L Glu70	102.72	ND5 His328	16.07	ND5 Cys279&NDUFB8 Cys115	17.57
270	SER	ND5		2	ND4L Glu70	100.81	ND5 TMH7 Gly222-Pro258	19.34	ND5 Cys279&NDUFB8 Cys115	14.86
283	ILE	ND5		2	ND4L Glu70	104.21	ND5 TMH7 Gly222-Pro265	4.40	ND5 Cys279&NDUFB8 Cys115	6.13
410	SER	ND5		2	ND4L Glu70	114.74	ND5 Lys392	12.77	ND5 Cys279&NDUFB8 Cys115	10.91
432	THR	ND5		2	ND4L Glu70	124.24	ND5 TMH12 Gly382-Tyr390	20.87	NDUFAB1 (β)-acyl chain	22.17
439	THR	ND5		2	ND4L Glu70	121.56	ND5 Lys336	33.58	NDUFB9 LYK (18-20)	18.56
449	THR	ND5		2	ND4L Glu70	116.00	ND5 Lys336	19.36	NDUFB9 LYK (18-20)	38.31
459	ALA	ND5		2	ND4L Glu70	121.84	ND5 TMH12 Gly382-Tyr390	9.46	NDUFB7 Cys68&Cys79	24.84
469	THR	ND5		2	ND4L Glu70	120.30	ND5 TMH12 Gly382-Tyr390	6.05	NDUFB7 Cys68&Cys79	13.51
471	ASN	ND5		2	ND4L Glu70	121.21	ND5 TMH12 Gly382-Tyr390	11.19	NDUFB7 Cys68&Cys79	10.56
478	PHE	ND5		2	ND4L Glu70	121.28	ND5 TMH12 Gly382-Tyr390	10.11	NDUFB7 Cys58-89	11.69
489	THR	ND5		2	ND4L Glu70	124.84	ND5 TMH12 Gly382-Tyr390	7.32	ND5 Cys279&NDUFB8 Cys115	21.91
495	PHE	ND5		2	ND4L Glu70	125.55	ND5 TMH12 Gly382-Tyr390	8.78	ND5 Cys279&NDUFB8 Cys115	25.76
518	CYS	ND5		2	ND4L Glu70	108.82	ND5 His230	18.02	NDUFAB1 (β)-acyl chain	17.74
530	PRO	ND5		2	ND4L Glu70	94.52	ND5 TMH7 Gly222-Pro288	5.92	ND5 Cys279&NDUFB8 Cys115	21.56
536	THR	ND5		2	ND4L Glu70	85.43	ND5 TMH7 Gly222-Pro289	13.51	ND5 Cys279&NDUFB8 Cys115	26.29
538	PRO	ND5		2	ND4L Glu70	83.04	ND5 TMH7 Gly222-Pro289	13.18	ND5 Cys279&NDUFB8 Cys115	27.98
541	GLY	ND5		2	ND4L Glu70	78.47	ND4 TMH7 Met207-Leu214	14.75	ND5 Cys279&NDUFB8 Cys115	29.23
549	PRO	ND5		2	ND4L Glu70	67.28	ND4 TMH7 Met207-Leu214	4.62	NDUFA11 Cys94&Cys114	23.85
556	THR	ND5		2	ND4L Glu70	58.45	ND5 Asp554 pi bulge	5.89	NDUFA11 Cys94&Cys114	19.83
573	THR	ND5		2	ND4L Glu70	36.53	ND2 His112	13.34	NDUFA11 Cys94&Cys114	18.99
575	ILE	ND5		2	ND4L Glu70	34.83	ND2 His112	12.61	NDUFA11 Cys94&Cys114	23.95
592	PHE	ND5		2	ND4L Glu34	21.60	ND2 TMH4 Gly107-Phe111	8.39	NDUFA11 Cys94&Cys114	32.91
598	THR	ND5		2	ND4L Glu70	24.66	ND2 TMH4 Gly107-Phe111	10.76	NDUFA11 Cys17-Cy74	30.66
14	MET	ND6		2	ND4L Glu34	16.24	ND2 Glu34	25.38	NDUFS5 Cys32&Cys65	32.17
31	VAL	ND6		2	ND4L Glu34	11.31	ND2 Glu34	23.39	NDUFS5 Cys32&Cys65	36.58
41	VAL	ND6		2	ND6 pi-bulge (Gly61-62)	17.48	ND2 Glu34	25.36	NDUFS5 Cys32&Cys65	25.99
86	VAL	ND6		2	ND4L Glu34	24.00	ND2 TMH4 Gly107-Phe111	26.65	NDUFAB1 (α)-acyl chain	35.93
97	ALA	ND6		2	ND4L Glu34	23.24	ND2 TMH4 Gly107-Phe111	24.96	NDUFS5 Cys32&Cys65	34.49
120	SER	ND6		2	ND4L Glu34	33.24	ND2 Glu34	36.53	NDUFS5 Cys32&Cys65	18.55
139	PRO	ND6		2	ND3 Glu68	16.65	ND2 Glu34	22.96	NDUFS5 Cys32&Cys65	21.04
156	THR	ND6		2	ND3 Glu68	8.00	ND2 Glu34	10.36	NDUFS5 Cys32&Cys65	29.64

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
159	THR	ND6		2	ND3 Glu68	8.61	ND2 Glu34	9.20	NDUFS5 Cys32&Cys65	33.64
165	TYR	ND6		2	ND3 Glu68	12.09	ND2 Glu34	15.48	NDUFS5 Cys32&Cys65	43.17
30	ASN	NDUFA10		2	NDUFS2 H59	58.47	ND2 Leu131 pi bulge	51.77	NDUFA10 ADP	6.62
13	LEU	NDUFA2		2	Fe-S cluster	67.52	ND2 His112	109.83	NDUFAB1 (α)-acyl chain	50.00
20	VAL	NDUFA3		2	ND3 Glu68	20.89	ND2 Glu34	37.39	NDUFA8 Cys87&Cys99	46.22
107	HIS	NDUFA6		2	ND3-Cys39	25.20	ND2 TMH4 Gly107-Phe111	66.74	NDUFAB1 (α)-acyl chain	11.13
65	PRO	NDUFA7		2	Fe-S cluster	33.96	ND2 Glu34	93.94	NDUFAB1 (α)-acyl chain	50.29
18	MET	NDUFB1		2	ND4L Glu70	65.55	ND2 TMH9 Gly252-Gly259	26.35	NDUFB10 Cys70&Cys77	40.91
141	ALA	NDUFB10		2	ND4L Glu70	69.80	ND4 TMH7 Met207-Leu214	16.58	NDUFB11 Cys112&NDUFB10 Cys148	12.54
52	PRO	NDUFB2		2	ND4L Glu70	133.75	ND5 TMH12 Gly382-Tyr390	14.29	NDUFB7 Cys58-89	13.68
62	LEU	NDUFB7		2	ND4L Glu70	126.05	ND5 TMH12 Gly382-Tyr390	19.68	NDUFB7 Cys58-89	5.89
96	SER	NDUFB8		2	ND4L Glu70	97.54	ND5 His230	19.33	NDUFAB1 (β)-acyl chain	17.48
120	VAL	NDUFB8		2	ND4L Glu70	110.70	ND5 TMH7 Gly222-Pro264	20.20	ND5 Cys279&NDUFB8 Cys115	9.25
123	VAL	NDUFS3		2	NDUFS2 H59	30.37	ND2 Leu131 pi bulge	63.57	NDUFA6 LYR (34-36)	29.01
8	VAL	NDUFV2		2	Fe-S cluster	16.31	ND2 Glu34	130.26	NDUFS6 Zinc	24.07
5	ASN	ND1		3	Q-site	21.97	ND2 Glu34	46.65	NDUFA8 Cys87&Cys99	26.48
15	ILE	ND1		3	Q-site	7.68	ND2 Glu34	50.35	NDUFA8 Cys87&Cys99	37.54
17	MET	ND1		3	Q-site	6.54	ND2 Glu34	46.90	NDUFA8 Cys87&Cys99	39.17
21	MET	ND1		3	Q-site	6.47	ND2 Glu34	49.78	NDUFA9 NADPH	37.75
27	ILE	ND1		3	Q-site	12.04	ND2 Glu34	54.79	NDUFA9 NADPH	45.62
68	ILE	ND1		3	ND1 Glu227	20.09	ND2 Glu34	38.16	NDUFA9 NADPH	29.64
76	THR	ND1		3	ND1 Glu227	12.49	ND2 Glu34	36.89	NDUFA9 NADPH	35.01
77	LEU	ND1		3	ND1 Glu227	12.43	ND2 Glu34	33.23	NDUFA9 NADPH	38.48
96	VAL	ND1		3	ND1 Glu192	20.77	ND2 Glu34	37.42	NDUFA8 Cys87&Cys99	23.94
112	ALA	ND1		3	ND1 Glu143	7.41	ND2 Glu34	29.31	NDUFS5 Cys32&Cys65	42.07
116	ILE	ND1		3	ND1 Glu227	8.75	ND2 Glu34	31.59	NDUFAB1 (α)-acyl chain	39.96
139	THR	ND1		3	ND1 Glu143	4.65	ND2 Glu34	26.99	NDUFS5 Cys32&Cys65	45.08
152	SER	ND1		3	ND1 Glu143	14.21	ND2 Glu34	29.41	NDUFA8 Cys87&Cys99	32.32
153	THR	ND1		3	ND1 Glu143	15.28	ND2 Glu34	32.45	NDUFA8 Cys87&Cys99	29.53
168	THR	ND1		3	ND1 Glu192	23.42	ND2 Glu34	40.85	NDUFA8 Cys87&Cys99	18.33
171	HIS	ND1		3	ND1 Glu192	27.84	ND2 Glu34	48.04	NDUFA8 Cys87&Cys99	17.94
196	THR	ND1		3	ND1 Glu192	8.59	ND2 Glu34	41.09	NDUFA8 Cys87&Cys99	46.31
213	ILE	ND1		3	ND3-Cys39	10.05	ND2 Glu34	42.92	NDUFA9 NADPH	30.90
239	THR	ND1		3	ND1 Glu192	14.27	ND2 Glu34	42.09	NDUFA8 Cys87&Cys99	27.27
241	ILE	ND1		3	ND1 Glu192	14.16	ND2 Glu34	37.76	NDUFA8 Cys87&Cys99	26.06
249	ALA	ND1		3	ND1 Glu192	29.57	ND2 Glu34	50.32	NDUFA8 Cys87&Cys99	10.69
258	TYR	ND1		3	ND1 Glu192	22.85	ND2 Glu34	50.95	NDUFA8 Cys87&Cys99	23.86

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
260	VAL	ND1		3	ND1 Glu192	19.26	ND2 Glu34	48.91	NDUFA8 Cys87&Cys99	27.39
265	LEU	ND1		3	ND1 Glu192	17.78	ND2 Glu34	49.21	NDUFA8 Cys87&Cys99	33.99
266	LEU	ND1		3	ND1 Glu192	14.20	ND2 Glu34	45.61	NDUFA8 Cys87&Cys99	34.52
268	SER	ND1		3	Q-site	15.40	ND2 Glu34	48.96	NDUFA8 Cys87&Cys99	38.65
273	ILE	ND1		3	Q-site	13.67	ND2 Glu34	46.67	NDUFA8 Cys87&Cys99	44.62
292	ASN	ND1		3	ND1 Glu143	14.63	ND2 Glu34	32.54	NDUFA8 Cys87&Cys99	48.07
300	LEU	ND1		3	ND3 Glu68	14.91	ND2 Glu34	31.17	NDUFA8 Cys87&Cys99	37.59
305	VAL	ND1		3	ND3 Glu68	14.84	ND2 Glu34	29.50	NDUFA8 Cys87&Cys99	33.75
310	THR	ND1		3	ND3 Glu68	22.85	ND2 Glu34	36.94	NDUFA8 Cys87&Cys99	32.73
313	SER	ND1		3	ND3 Glu68	24.96	ND2 Glu34	38.29	NDUFA8 Cys87&Cys99	25.05
318	THR	ND1		3	ND3 Glu68	26.79	ND2 Glu34	36.46	NDUFA8 Cys87&Cys99	17.89
11	SER	ND2		3	ND4L Glu70	18.10	ND2 Glu34	14.22	NDUFS5 Cys32&Cys65	40.47
15	ALA	ND2		3	ND4L Glu70	15.57	ND2 Glu34	10.82	NDUFS5 Cys32&Cys65	34.61
19	ILE	ND2		3	ND4L Glu70	14.90	ND2 Glu34	10.20	NDUFS5 Cys32&Cys65	28.81
48	MET	ND2		3	ND4L Glu70	19.79	ND2 Leu131 pi bulge	20.31	NDUFA10 ADP	34.93
49	ASN	ND2		3	ND4L Glu70	22.07	ND2 Leu131 pi bulge	19.81	NDUFA10 ADP	32.89
62	THR	ND2		3	ND4L Glu70	10.12	ND2 Lys105	7.40	NDUFS5 Cys32&Cys65	36.96
69	ILE	ND2		3	ND4L Glu70	10.42	ND2 Lys105	9.04	NDUFS5 Cys32&Cys65	26.27
78	ASN	ND2		3	ND4L Glu70	20.16	ND2 Glu34	17.34	NDUFS5 Cys32&Cys65	12.01
79	MET	ND2		3	ND4L Glu70	21.16	ND2 Glu34	18.84	NDUFS5 Cys32&Cys65	11.74
81	SER	ND2		3	ND4L Glu70	25.52	ND2 Glu34	22.73	NDUFS5 Cys32&Cys65	10.32
85	THR	ND2		3	ND4L Glu70	21.32	ND2 Glu34	18.41	NDUFS5 Cys32&Cys65	15.83
94	SER	ND2		3	ND4L Glu70	19.10	ND2 Lys105	14.79	NDUFS5 Cys32&Cys65	20.80
98	ILE	ND2		3	ND4L Glu70	15.26	ND2 Lys105	8.97	NDUFS5 Cys32&Cys65	25.33
115	VAL	ND2		3	ND4L Glu70	18.57	ND2 His112	5.60	NDUFA11 Cys94&Cys114	37.31
125	THR	ND2		3	ND4L Glu70	23.37	ND2 Leu131 pi bulge	9.85	NDUFA10 ADP	37.32
139	ILE	ND2		3	ND4L Glu70	17.82	ND2 Lys135	6.00	NDUFS5 Cys42&Cys55	29.45
141	ILE	ND2		3	ND4L Glu70	16.62	ND2 Lys135	10.04	NDUFS5 Cys42&Cys55	25.50
145	ILE	ND2		3	ND4L Glu70	21.54	ND2 Lys105	14.59	NDUFS5 Cys42&Cys55	20.09
147	PRO	ND2		3	ND4L Glu70	26.70	ND2 TMH4 Gly107-Phe111	17.03	NDUFS5 Cys42&Cys55	18.71
148	SER	ND2		3	ND4L Glu70	26.19	ND2 TMH4 Gly107-Phe111	17.76	NDUFS5 Cys42&Cys55	17.62
156	THR	ND2		3	ND4L Glu70	27.41	ND2 TMH4 Gly107-Phe111	6.88	NDUFA11 Cys94&Cys114	26.21
159	ILE	ND2		3	ND4L Glu70	28.55	ND2 TMH4 Gly107-Phe111	7.21	NDUFA11 Cys94&Cys114	24.91
166	SER	ND2		3	ND4L Glu70	30.69	ND2 His112	9.48	NDUFA11 Cys94&Cys114	25.62
187	MET	ND2		3	ND4L Glu70	22.75	ND2 His186	3.78	NDUFA11 Cys94&Cys114	34.24
200	MET	ND2		3	ND4L Glu70	29.75	ND2 Lys263	9.82	NDUFS5 Cys42&Cys55	30.66
209	ILE	ND2		3	ND4L Glu70	25.37	ND2 Leu131 pi bulge	7.45	NDUFS5 Cys42&Cys55	41.02
219	LEU	ND2		3	ND4L Glu70	31.89	ND2 Leu131 pi bulge	12.33	NDUFA10 ADP	33.21
220	ASN	ND2		3	ND4L Glu70	31.73	ND2 Leu131 pi bulge	14.04	NDUFA10 ADP	32.38
222	ASN	ND2		3	ND4L Glu70	36.11	ND2 Leu131 pi bulge	17.69	NDUFA10 ADP	28.08

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
238	THR	ND2		3	ND4L Glu70	47.09	ND2 TMH9 Gly252-Gly259	19.32	NDUFA10 ADP	31.12
239	TRP	ND2		3	ND4L Glu70	46.03	ND2 TMH9 Gly252-Gly259	16.48	NDUFA10 ADP	34.51
241	THR	ND2		3	ND4L Glu70	43.66	ND2 TMH9 Gly252-Gly259	13.68	NDUFA10 ADP	35.18
242	PRO	ND2		3	ND4L Glu70	43.48	ND2 TMH9 Gly252-Gly259	10.12	NDUFA10 ADP	39.00
246	SER	ND2		3	ND4L Glu70	39.11	ND2 TMH9 Gly252-Gly259	5.26	NDUFA11 Cys94&Cys114	40.42
266	ILE	ND2		3	ND4L Glu70	33.78	ND2 Lys263	5.40	NDUFA11 Cys94&Cys114	33.60
267	ILE	ND2		3	ND4L Glu70	36.84	ND2 Lys263	6.58	NDUFA11 Cys94&Cys114	31.01
276	LEU	ND2		3	ND4L Glu70	40.25	ND2 Lys263	14.47	NDUFA11 Cys94&Cys114	25.39
277	ILE	ND2		3	ND4L Glu70	39.20	ND2 Lys263	15.25	NDUFA11 Cys94&Cys114	22.12
281	ILE	ND2		3	ND4L Glu70	36.48	ND2 Lys263	12.55	NDUFA11 Cys94&Cys114	20.71
285	ILE	ND2		3	ND4L Glu70	35.39	ND2 Lys263	10.16	NDUFA11 Cys94&Cys114	21.45
286	THR	ND2		3	ND4L Glu70	34.48	ND2 Lys263	6.85	NDUFA11 Cys94&Cys114	24.98
317	PHE	ND2		3	ND4L Glu70	45.47	ND2 Leu131 pi bulge	29.29	NDUFA10 ADP	23.46
320	THR	ND2		3	ND4L Glu70	38.85	ND2 Leu131 pi bulge	23.48	NDUFA10 ADP	29.95
330	ILE	ND2		3	ND4L Glu70	36.90	ND2 TMH9 Gly252-Gly259	8.03	NDUFA10 ADP	43.93
332	LEU	ND2		3	ND4L Glu70	36.49	ND2 TMH9 Gly252-Gly259	5.50	NDUFB10 Cys70&Cys77	44.54
333	THR	ND2		3	ND4L Glu70	36.18	ND2 TMH9 Gly252-Gly259	3.32	NDUFA11 Cys94&Cys114	44.25
343	MET	ND2		3	ND4L Glu70	36.26	ND2 TMH9 Gly252-Gly259	8.86	NDUFB10 Cys70&Cys77	36.79
345	MET	ND2		3	ND4L Glu70	35.21	ND2 TMH9 Gly252-Gly259	14.81	NDUFS5 Cys42&Cys55	32.26
346	ILE	ND2		3	ND4L Glu70	31.98	ND2 TMH9 Gly252-Gly259	14.86	NDUFS5 Cys42&Cys55	31.49
14	ALA	ND3		3	Q-site	14.72	ND2 Glu34	41.28	NDUFA9 NADPH	32.76
19	ILE	ND3		3	Q-site	18.93	ND2 Glu34	45.37	NDUFA9 NADPH	25.95
24	LEU	ND3		3	Q-site	18.06	ND2 Glu34	54.28	NDUFA9 NADPH	17.52
44	MET	ND3		3	ND3-Cys39	10.72	ND2 Glu34	44.75	NDUFAB1 (α)-acyl chain	17.46
89	MET	ND3		3	ND3 Glu68	13.07	ND2 Glu34	25.17	NDUFS5 Cys32&Cys65	33.75
90	SER	ND3		3	ND3 Glu68	10.42	ND2 Glu34	21.57	NDUFS5 Cys32&Cys65	31.31
93	LEU	ND3		3	ND3 Glu68	7.59	ND2 Glu34	21.25	NDUFS5 Cys32&Cys65	35.44
97	ILE	ND3		3	ND3 Glu68	4.68	ND2 Glu34	19.17	NDUFS5 Cys32&Cys65	38.43
105	GLU	ND3		3	ND3 Glu68	13.03	ND2 Glu34	23.28	NDUFS5 Cys32&Cys65	48.28
6	VAL	ND4		3	ND4L Glu70	59.27	ND2 TMH9 Gly252-Gly259	17.73	NDUFB10 Cys70&Cys77	35.02
21	HIS	ND4		3	ND4L Glu70	68.40	ND4 Ala223 pi bulge	25.71	NDUFA10 ADP	45.01
29	THR	ND4		3	ND4L Glu70	70.25	ND4 Lys237	23.75	NDUFB10 Cys70&Cys77	47.33
45	ILE	ND4		3	ND4L Glu70	74.41	ND4 Lys237	23.58	NDUFB10 Cys70&Cys77	20.74
46	ASN	ND4		3	ND4L Glu70	73.00	ND4 Lys237	21.80	NDUFB11 Cys112&NDUFB10 Cys148	17.84
52	CYS	ND4		3	ND4L Glu70	63.89	ND4 Lys237	23.87	NDUFB10 Cys70&Cys77	19.40
55	THR	ND4		3	ND4L Glu70	55.45	ND2 TMH9 Gly252-Gly259	17.56	NDUFB10 Cys70&Cys77	22.12
58	SER	ND4		3	ND4L Glu70	63.86	ND4 Lys237	16.69	NDUFB10 Cys70&Cys77	22.75
70	THR	ND4		3	ND4L Glu70	65.49	ND4 Lys237	13.13	NDUFB10 Cys70&Cys77	37.73
85	SER	ND4		3	ND4L Glu70	66.13	ND4 Ala223 pi bulge	17.53	NDUFA10 ADP	41.40

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
104	ILE	ND4		3	ND4L Glu70	56.00	ND2 TMH9 Gly252-Gly259	11.53	NDUFB10 Cys70&Cys77	36.19
107	ILE	ND4		3	ND4L Glu70	58.39	ND4 Glu123	13.05	NDUFB10 Cys70&Cys77	31.91
121	PHE	ND4		3	ND4L Glu70	49.93	ND4 Glu123	5.35	NDUFB10 Cys70&Cys77	32.30
129	THR	ND4		3	ND4L Glu70	52.34	ND4 Lys206	8.71	NDUFA11 Cys94&Cys114	36.65
166	TYR	ND4		3	ND4L Glu70	45.79	ND2 Lys263	15.27	NDUFB10 Cys70&Cys77	26.90
170	THR	ND4		3	ND4L Glu70	49.11	ND2 Lys263	20.24	NDUFB10 Cys70&Cys77	21.86
187	SER	ND4		3	ND4L Glu70	58.05	ND4 TMH7 Met207-Leu214	23.66	NDUFB11 Cys112&NDUFB10 Cys148	20.03
193	ASN	ND4		3	ND4L Glu70	54.21	ND4 TMH7 Met207-Leu214	15.94	NDUFA11 Cys94&Cys114	21.18
201	MET	ND4		3	ND4L Glu70	52.92	ND4 TMH7 Met207-Leu214	5.97	NDUFA11 Cys94&Cys114	22.36
225	ILE	ND4		3	ND4L Glu70	63.28	ND4 Ala223 pi bulge	5.63	NDUFA11 Cys94&Cys114	43.44
232	ALA	ND4		3	ND4L Glu70	62.86	ND4 His220	3.58	NDUFA11 Cys94&Cys114	36.91
291	ILE	ND4		3	ND4L Glu70	64.86	ND4 TMH7 Met207-Leu214	4.49	NDUFA11 Cys94&Cys114	30.05
317	ILE	ND4		3	ND4L Glu70	69.63	ND4 His319	5.52	NDUFB11 Cys112&NDUFB10 Cys148	31.34
320	GLY	ND4		3	ND4L Glu70	67.88	ND4 His319	3.78	NDUFB11 Cys112&NDUFB10 Cys148	35.96
350	THR	ND4		3	ND4L Glu70	88.82	ND4 TMH12 Leu367-Thr372	22.00	NDUFB9 LYK (18-20)	34.11
365	ALA	ND4		3	ND4L Glu70	78.32	ND4 TMH12 Leu367-Thr372	5.21	ND5 Cys279&NDUFB8 Cys115	37.97
391	ILE	ND4		3	ND4L Glu70	78.20	ND4 Glu378	16.53	ND5 Cys279&NDUFB8 Cys115	26.24
418	SER	ND4		3	ND4L Glu70	87.66	ND4 TMH12 Leu367-Thr372	22.90	NDUFB9 LYK (18-20)	26.32
419	LEU	ND4		3	ND4L Glu70	86.48	ND4 TMH12 Leu367-Thr372	23.71	NDUFB9 LYK (18-20)	28.91
420	THR	ND4		3	ND4L Glu70	83.56	ND4 TMH12 Leu367-Thr372	25.13	NDUFB9 LYK (18-20)	30.79
421	HIS	ND4		3	ND4L Glu70	84.86	ND4 TMH12 Leu367-Thr372	27.86	NDUFB9 LYK (18-20)	31.67
425	ASN	ND4		3	ND4L Glu70	78.29	ND4 Ala223 pi bulge	22.28	NDUFB9 LYK (18-20)	41.01
434	ASN	ND4		3	ND4L Glu70	78.54	ND4 TMH12 Leu367-Thr372	19.70	NDUFB9 LYK (18-20)	50.46
435	THR	ND4		3	ND4L Glu70	77.56	ND4 TMH12 Leu367-Thr372	18.93	NDUFB11 Cys112&NDUFB10 Cys148	50.73
455	THR	ND4		3	ND4L Glu70	74.16	ND4 Glu378	13.94	NDUFB11 Cys112&NDUFB10 Cys148	22.87
21	VAL	ND4L		3	ND4L Glu34	15.93	ND2 TMH4 Gly107-Phe111	13.34	NDUFS5 Cys32&Cys65	40.52
59	VAL	ND4L		3	ND4L Glu70	14.99	ND2 Glu34	11.93	NDUFS5 Cys32&Cys65	18.51
79	VAL	ND4L		3	ND4L Glu34	13.91	ND2 Glu34	20.97	NDUFA6 LYR (34-36)	38.94
80	SER	ND4L		3	ND4L Glu70	15.49	ND2 Glu34	21.16	NDUFA6 LYR (34-36)	38.34

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
2	THR	ND5		3	ND4L Glu70	109.80	ND4 TMH12 Leu367-Thr372	27.79	NDUFB10 Cys106&Cys118	20.27
14	SER	ND5		3	ND4L Glu70	104.05	ND4 TMH12 Leu367-Thr372	19.62	NDUFB10 Cys106&Cys118	31.87
18	PRO	ND5		3	ND4L Glu70	100.97	ND4 TMH12 Leu367-Thr372	19.09	NDUFB10 Cys106&Cys118	39.01
27	ASN	ND5		3	ND4L Glu70	99.99	ND4 TMH12 Leu367-Thr372	26.05	NDUFB9 LYK (18-20)	46.48
30	ASN	ND5		3	ND4L Glu70	105.77	ND5 His248	29.05	NDUFB9 LYK (18-20)	47.73
34	HIS	ND5		3	ND4L Glu70	109.31	ND5 His248	22.83	NDUFB7 Cys68&Cys79	41.26
38	SER	ND5		3	ND4L Glu70	110.72	ND5 Lys336	23.57	NDUFB7 Cys68&Cys79	36.73
39	ILE	ND5		3	ND4L Glu70	108.05	ND5 His248	22.93	NDUFB7 Cys68&Cys79	36.23
40	VAL	ND5		3	ND4L Glu70	108.04	ND5 His248	19.64	NDUFB7 Cys68&Cys79	33.96
45	ILE	ND5		3	ND4L Glu70	114.73	ND5 His332	21.10	NDUFB7 Cys68&Cys79	26.85
53	MET	ND5		3	ND4L Glu70	117.91	ND5 His328	18.95	NDUFB7 Cys68&Cys79	14.45
59	GLN	ND5		3	ND4L Glu70	110.95	ND5 His328	22.95	NDUFB10 Cys106&Cys118	14.83
63	ILE	ND5		3	ND4L Glu70	100.12	ND4 TMH12 Leu367-Thr372	18.91	NDUFB10 Cys106&Cys118	14.76
65	ASN	ND5		3	ND4L Glu70	95.78	ND4 TMH12 Leu367-Thr372	15.90	NDUFB10 Cys106&Cys118	18.33
67	HIS	ND5		3	ND4L Glu70	88.89	ND4 TMH12 Leu367-Thr372	12.50	NDUFB10 Cys106&Cys118	20.58
70	THR	ND5		3	ND4L Glu70	82.90	ND4 TMH12 Leu367-Thr372	15.38	NDUFB10 Cys106&Cys118	21.58
76	LEU	ND5		3	ND4L Glu70	86.94	ND4 TMH12 Leu367-Thr372	11.70	NDUFB10 Cys106&Cys118	19.39
88	MET	ND5		3	ND4L Glu70	110.82	ND5 His328	12.50	NDUFB7 Cys68&Cys79	21.43
93	ALA	ND5		3	ND4L Glu70	107.38	ND5 His332	10.59	NDUFB7 Cys68&Cys79	28.62
109	ASN	ND5		3	ND4L Glu70	107.80	ND5 His248	20.39	NDUFB9 LYK (18-20)	32.56
113	ASN	ND5		3	ND4L Glu70	98.54	ND5 His248	19.92	NDUFB9 LYK (18-20)	35.12
127	THR	ND5		3	ND4L Glu70	95.48	ND5 Glu145	8.54	NDUFB10 Cys106&Cys118	29.79
132	VAL	ND5		3	ND4L Glu70	100.26	ND5 Lys223	10.27	NDUFB10 Cys106&Cys118	23.37
141	PHE	ND5		3	ND4L Glu70	89.06	ND4 TMH12 Leu367-Thr372	4.43	NDUFB10 Cys106&Cys118	25.31
147	VAL	ND5		3	ND4L Glu70	93.54	ND5 Lys223	4.99	ND5 Cys279&NDUFB8 Cys115	27.42
162	ALA	ND5		3	ND4L Glu70	90.26	ND5 His248	16.13	NDUFB9 LYK (18-20)	25.28
169	ILE	ND5		3	ND4L Glu70	86.97	ND5 Arg176	9.51	ND5 Cys279&NDUFB8 Cys115	28.39
172	ILE	ND5		3	ND4L Glu70	87.83	ND5 Arg176	5.39	ND5 Cys279&NDUFB8 Cys115	25.38
177	ILE	ND5		3	ND4L Glu70	86.50	ND5 Arg176	3.79	ND5 Cys279&NDUFB8 Cys115	20.17
196	TRP	ND5		3	ND4L Glu70	89.41	ND4 TMH12 Leu367-Thr372	12.94	NDUFB10 Cys106&Cys118	16.07

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
206	ALA	ND5		3	ND4L Glu70	97.89	ND5 Asp179	22.77	NDUFB10 Cys106&Cys118	13.57
209	SER	ND5		3	ND4L Glu70	93.36	ND5 Asp179	18.31	ND5 Cys279&NDUFB8 Cys115	19.81
211	THR	ND5		3	ND4L Glu70	89.69	ND5 Asp179	14.21	ND5 Cys279&NDUFB8 Cys115	18.02
271	PRO	ND5		3	ND4L Glu70	103.33	ND5 TMH7 Gly222-Pro264	17.74	ND5 Cys279&NDUFB8 Cys115	11.89
273	ILE	ND5		3	ND4L Glu70	99.56	ND5 TMH7 Gly222-Pro258	13.53	ND5 Cys279&NDUFB8 Cys115	9.62
290	VAL	ND5		3	ND4L Glu70	108.30	ND5 His230	10.67	ND5 Cys279&NDUFB8 Cys115	16.34
304	PHE	ND5		3	ND4L Glu70	105.76	ND5 Lys336	8.28	ND5 Cys279&NDUFB8 Cys115	18.44
317	ILE	ND5		3	ND4L Glu70	106.10	ND5 His328	8.49	ND5 Cys279&NDUFB8 Cys115	11.65
323	HIS	ND5		3	ND4L Glu70	110.82	ND5 His328	8.44	ND5 Cys279&NDUFB8 Cys115	20.01
331	THR	ND5		3	ND4L Glu70	114.30	ND5 His332	3.78	NDUFB7 Cys68&Cys79	22.49
377	SER	ND5		3	ND4L Glu70	118.45	ND5 TMH12 Gly382-Tyr390	6.96	ND5 Cys279&NDUFB8 Cys115	26.73
378	LEU	ND5		3	ND4L Glu70	121.49	ND5 TMH12 Gly382-Tyr390	4.24	ND5 Cys279&NDUFB8 Cys115	27.42
381	ALA	ND5		3	ND4L Glu70	119.78	ND5 TMH12 Gly382-Tyr390	3.77	ND5 Cys279&NDUFB8 Cys115	23.27
400	ASN	ND5		3	ND4L Glu70	117.57	ND5 TMH12 Gly382-Tyr390	14.95	ND5 Cys279&NDUFB8 Cys115	16.05
403	TYR	ND5		3	ND4L Glu70	112.13	ND5 TMH7 Gly222-Pro264	20.13	ND5 Cys279&NDUFB8 Cys115	12.39
413	LEU	ND5		3	ND4L Glu70	115.92	ND5 Lys392	8.26	ND5 Cys279&NDUFB8 Cys115	13.53
420	SER	ND5		3	ND4L Glu70	118.30	ND5 TMH12 Gly382-Tyr390	8.28	ND5 Cys279&NDUFB8 Cys115	21.86
423	SER	ND5		3	ND4L Glu70	116.57	ND5 Lys336	10.85	ND5 Cys279&NDUFB8 Cys115	24.31
424	THR	ND5		3	ND4L Glu70	119.78	ND5 TMH12 Gly382-Tyr390	11.99	NDUFAB1 (β)-acyl chain	26.27
440	LEU	ND5		3	ND4L Glu70	118.65	ND5 Lys336	32.38	NDUFB9 LYK (18-20)	18.59
441	THR	ND5		3	ND4L Glu70	118.95	ND5 Lys336	29.01	NDUFB9 LYK (18-20)	20.44
442	ASN	ND5		3	ND4L Glu70	119.08	ND5 Lys336	27.51	NDUFB9 LYK (18-20)	23.80
447	ASN	ND5		3	ND4L Glu70	119.48	ND5 Lys336	22.28	NDUFB9 LYK (18-20)	36.35
456	ARG	ND5		3	ND4L Glu70	119.09	ND5 TMH12 Gly382-Tyr390	13.61	NDUFB7 Cys68&Cys79	28.10
461	SER	ND5		3	ND4L Glu70	118.81	ND5 TMH12 Gly382-Tyr390	5.08	NDUFB7 Cys68&Cys79	24.71
472	ILE	ND5		3	ND4L Glu70	118.10	ND5 TMH12 Gly382-Tyr390	11.48	NDUFB7 Cys68&Cys79	12.88
473	SER	ND5		3	ND4L Glu70	118.40	ND5 TMH12 Gly382-Tyr390	12.82	NDUFB7 Cys68&Cys79	11.81
476	SER	ND5		3	ND4L Glu70	116.04	ND5 TMH12 Gly382-Tyr390	13.72	NDUFB7 Cys58-89	14.40
477	PRO	ND5		3	ND4L Glu70	119.57	ND5 TMH12 Gly382-Tyr390	11.56	NDUFB7 Cys58-89	12.51

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
491	LEU	ND5		3	ND4L Glu70	124.58	ND5 TMH12 Gly382-Tyr390	3.64	ND5 Cys279&NDUFB8 Cys115	23.18
492	ALA	ND5		3	ND4L Glu70	126.16	ND5 TMH12 Gly382-Tyr390	7.39	ND5 Cys279&NDUFB8 Cys115	24.37
493	VAL	ND5		3	ND4L Glu70	122.69	ND5 TMH12 Gly382-Tyr390	7.83	ND5 Cys279&NDUFB8 Cys115	21.26
500	THR	ND5		3	ND4L Glu70	122.96	ND5 TMH12 Gly382-Tyr390	16.56	NDUFAB1 (β)-acyl chain	25.47
508	THR	ND5		3	ND4L Glu70	124.95	ND5 TMH12 Gly382-Tyr390	25.71	NDUFAB1 (β)-acyl chain	18.51
509	ASN	ND5		3	ND4L Glu70	122.35	ND5 TMH12 Gly382-Tyr390	27.38	NDUFAB1 (β)-acyl chain	14.76
513	MET	ND5		3	ND4L Glu70	119.60	ND5 His230	29.38	NDUFAB1 (β)-acyl chain	9.96
517	LEU	ND5		3	ND4L Glu70	111.43	ND5 His230	20.98	NDUFAB1 (β)-acyl chain	15.03
524	ASN	ND5		3	ND4L Glu70	105.79	ND5 His230	12.68	NDUFAB1 (β)-acyl chain	15.82
525	MET	ND5		3	ND4L Glu70	103.54	ND5 His230	9.81	NDUFAB1 (β)-acyl chain	18.66
546	GLN	ND5		3	ND4L Glu70	71.22	ND5 Asp554 pi bulge	7.93	NDUFA11 Cys94&Cys114	28.93
564	LYS	ND5		3	ND4L Glu70	48.60	ND4 TMH7 Met207-Leu214	8.81	NDUFA11 Cys94&Cys114	17.50
565	THR	ND5		3	ND4L Glu70	47.70	ND4 TMH7 Met207-Leu214	11.30	NDUFA11 Cys94&Cys114	14.06
569	HIS	ND5		3	ND4L Glu70	42.32	ND4 TMH7 Met207-Leu214	16.05	NDUFA11 Cys94&Cys114	15.29
571	ILE	ND5		3	ND4L Glu70	39.41	ND2 His112	16.47	NDUFA11 Cys94&Cys114	20.52
577	THR	ND5		3	ND4L Glu70	29.90	ND2 His112	7.22	NDUFA11 Cys94&Cys114	24.74
579	THR	ND5		3	ND4L Glu70	30.70	ND2 His112	10.97	NDUFA11 Cys94&Cys114	28.93
596	ILE	ND5		3	ND4L Glu34	25.52	ND2 TMH4 Gly107-Phe111	11.47	NDUFA11 Cys17-Cy74	29.73
602	ILE	ND5		3	ND4L Glu70	29.90	ND2 TMH4 Gly107-Phe111	16.80	NDUFS5 Cys42&Cys55	26.76
603	THR	ND5		3	ND4L Glu70	33.20	ND2 TMH4 Gly107-Phe111	18.91	NDUFA11 Cys17-Cy74	26.34
2	MET	ND6		3	ND4L Glu34	31.91	ND2 Glu34	37.92	NDUFS5 Cys32&Cys65	26.40
4	ALA	ND6		3	ND4L Glu34	26.15	ND2 Glu34	32.66	NDUFS5 Cys32&Cys65	25.76
34	VAL	ND6		3	ND4L Glu34	13.03	ND2 Glu34	24.31	NDUFS5 Cys32&Cys65	34.24
35	SER	ND6		3	ND4L Glu34	11.71	ND2 Glu34	21.78	NDUFS5 Cys32&Cys65	31.01
37	VAL	ND6		3	ND6 pi-bulge (Gly61-62)	12.81	ND2 Glu34	23.45	NDUFS5 Cys32&Cys65	30.54
38	VAL	ND6		3	ND4L Glu34	15.27	ND2 Glu34	23.83	NDUFS5 Cys32&Cys65	28.50
42	ILE	ND6		3	ND4L Glu34	18.12	ND2 Glu34	23.93	NDUFS5 Cys32&Cys65	22.72
45	ASN	ND6		3	ND6 pi-bulge (Gly61-62)	22.38	ND2 Glu34	27.49	NDUFS5 Cys32&Cys65	21.38
54	MET	ND6		3	ND3 Asp66	10.28	ND2 Glu34	17.95	NDUFS5 Cys32&Cys65	27.00
81	ALA	ND6		3	ND4L Glu34	17.96	ND2 TMH4 Gly107-Phe111	28.59	NDUFAB1 (α)-acyl chain	33.24
85	GLY	ND6		3	ND4L Glu34	22.55	ND2 TMH4 Gly107-Phe111	28.16	NDUFAB1 (α)-acyl chain	33.81
90	VAL	ND6		3	ND4L Glu34	22.09	ND2 TMH4 Gly107-Phe111	23.67	NDUFAB1 (α)-acyl chain	42.49
94	VAL	ND6		3	ND4L Glu34	20.87	ND2 TMH4 Gly107-Phe111	21.74	NDUFS5 Cys32&Cys65	36.85
102	LEU	ND6		3	ND4L Glu34	23.93	ND2 TMH4 Gly107-Phe111	25.02	NDUFS5 Cys32&Cys65	26.33
106	VAL	ND6		3	ND4L Glu34	28.78	ND2 TMH4 Gly107-Phe111	28.68	NDUFS5 Cys32&Cys65	23.75

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
121	VAL	ND6		3	ND4L Glu34	29.75	ND2 Glu34	33.67	NDUFS5 Cys32&Cys65	19.12
134	LEU	ND6		3	ND3 Glu68	28.89	ND2 Glu34	34.45	NDUFS5 Cys32&Cys65	19.84
135	ILE	ND6		3	ND3 Glu68	25.56	ND2 Glu34	31.01	NDUFS5 Cys32&Cys65	18.59
140	ILE	ND6		3	ND3 Glu68	18.35	ND2 Glu34	24.22	NDUFS5 Cys32&Cys65	19.60
142	ALA	ND6		3	ND3 Glu68	13.75	ND2 Glu34	19.02	NDUFS5 Cys32&Cys65	20.94
150	ARG	ND6		3	ND3 Glu68	13.24	ND2 Glu34	18.39	NDUFS5 Cys32&Cys65	26.10
155	VAL	ND6		3	ND3 Glu68	10.32	ND2 Glu34	11.31	NDUFS5 Cys32&Cys65	29.74
161	PHE	ND6		3	ND3 Glu68	6.52	ND2 Glu34	14.08	NDUFS5 Cys32&Cys65	38.16
162	VAL	ND6		3	ND3 Glu68	9.95	ND2 Glu34	12.04	NDUFS5 Cys32&Cys65	38.74
174	ASN	ND6		3	ND4L Glu70	22.87	ND2 Glu34	25.79	NDUFA10 ADP	34.81
2	ARG	NDUFA10		3	ND4L Glu70	37.15	ND2 Leu131 pi bulge	27.01	NDUFA10 ADP	34.68
137	ALA	NDUFA10		3	ND4L Glu70	56.79	ND2 Leu131 pi bulge	47.40	NDUFA10 ADP	9.82
142	GLY	NDUFA10		3	ND4L Glu70	61.51	ND4 Ala223 pi bulge	48.56	NDUFA10 ADP	15.88
165	PRO	NDUFA10		3	ND4L Glu70	40.18	ND2 Leu131 pi bulge	36.86	NDUFA10 ADP	18.91
182	ARG	NDUFA10		3	NDUFS2 H59	66.36	ND2 Leu131 pi bulge	61.10	NDUFA10 ADP	9.15
19	ARG	NDUFA11		3	ND4L Glu34	49.75	ND2 TMH4 Gly107-Phe111	31.76	NDUFA11 Cys17-Cy74	6.33
65	VAL	NDUFA11		3	ND4L Glu70	46.97	ND2 TMH4 Gly107-Phe111	26.36	NDUFA11 Cys94&Cys114	10.30
99	THR	NDUFA11		3	ND4L Glu70	50.78	ND4 TMH7 Met207-Leu214	28.01	NDUFA11 Cys94&Cys114	9.21
111	ALA	NDUFA11		3	ND4L Glu70	52.58	ND4 TMH7 Met207-Leu214	18.43	NDUFA11 Cys94&Cys114	4.42
128	ARG	NDUFA11		3	ND4L Glu70	46.60	ND2 TMH4 Gly107-Phe111	26.49	NDUFA11 Cys17-Cy74	16.52
60	ARG	NDUFA12		3	Fe-S cluster	29.51	ND2 Glu34	80.62	NDUFS6 Zinc	41.63
9	MET	NDUFA13		3	Fe-S cluster	31.26	ND2 Glu34	91.45	NDUFS6 Zinc	38.45
24	LEU	NDUFA13		3	NDUFS2 D160	25.85	ND2 Glu34	61.70	NDUFA8 Cys87&Cys99	58.87
41	THR	NDUFA13		3	ND1 Glu192	27.29	ND2 Glu34	46.90	NDUFA8 Cys87&Cys99	32.59
69	ALA	NDUFA13		3	ND3 Glu68	41.75	ND2 Glu34	47.82	NDUFA8 Cys35&Cys65	9.34
80	ARG	NDUFA13		3	ND3 Glu68	48.36	ND2 Glu34	49.22	NDUFA8 Cys45&Cys55	10.74
87	ARG	NDUFA13		3	ND4L Glu34	49.83	ND2 Glu34	50.06	NDUFA8 Cys45&Cys55	16.63
34	ASP	NDUFA2		3	Fe-S cluster	55.83	ND2 His112	120.22	NDUFAB1 (α)-acyl chain	53.79
62	PRO	NDUFA2		3	Fe-S cluster	54.20	ND2 His112	106.10	NDUFAB1 (α)-acyl chain	39.77
63	LYS	NDUFA2		3	Fe-S cluster	55.50	ND2 His112	105.64	NDUFAB1 (α)-acyl chain	39.81
93	VAL	NDUFA2		3	Fe-S cluster	65.73	ND2 His112	111.73	NDUFAB1 (α)-acyl chain	49.04
11	ALA	NDUFA3		3	ND3 Glu68	26.34	ND2 Glu34	41.04	NDUFA10 ADP	54.82
49	PRO	NDUFA3		3	ND3 Glu68	31.98	ND2 Glu34	38.90	NDUFA8 Cys35&Cys65	20.14
2	GLY	NDUFA5		3	NDUFS2 D160	46.83	ND2 Glu34	67.79	NDUFA10 ADP	39.85
63	ARG	NDUFA6		3	ND3-Cys39	32.63	ND2 His112	60.36	NDUFA6 LYR (34-36)	4.85
4	THR	NDUFA7		3	Q-site	31.83	ND2 Glu34	71.68	NDUFA8 Cys87&Cys99	49.35
7	ILE	NDUFA7		3	Q-site	28.58	ND2 Glu34	66.00	NDUFA8 Cys87&Cys99	46.60
29	GLU	NDUFA7		3	NDUFS2 Y108	31.92	ND2 Glu34	81.78	NDUFS6 Zinc	45.83
37	PRO	NDUFA7		3	Fe-S cluster	25.81	ND2 Glu34	96.57	NDUFS6 Zinc	26.93
50	ASN	NDUFA7		3	Fe-S cluster	23.64	ND2 Glu34	110.97	NDUFS6 Zinc	34.07

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
57	ASP	NDUFA7		3	Fe-S cluster	22.29	ND2 Glu34	99.87	NDUFS6 Zinc	36.37
59	ARG	NDUFA7		3	Fe-S cluster	23.16	ND2 Glu34	95.58	NDUFS6 Zinc	37.64
43	MET	NDUFA8		3	ND3 Glu68	44.13	ND2 Glu34	46.55	NDUFA8 Cys45&Cys55	5.21
54	ARG	NDUFA8		3	ND3 Glu68	50.81	ND2 Glu34	49.75	NDUFA8 Cys45&Cys55	3.82
134	ARG	NDUFA8		3	ND3 Glu68	54.98	ND2 Glu34	55.02	NDUFA8 Cys45&Cys55	8.42
21	ALA	NDUFA9		3	Fe-S cluster	35.76	ND2 Glu34	82.47	NDUFA9 NADPH	14.46
135	LEU	NDUFA9		3	ND3-Cys39	44.95	ND2 TMH4 Gly107-Phe111	71.76	NDUFA9 NADPH	7.64
187	ARG	NDUFA9		3	ND3-Cys39	31.50	ND2 TMH4 Gly107-Phe111	46.42	NDUFA9 NADPH	21.68
206	TYR	NDUFA9		3	ND3-Cys39	32.83	ND2 TMH4 Gly107-Phe111	64.00	NDUFA9 NADPH	9.18
235	ARG	NDUFA9		3	ND3-Cys39	43.50	ND2 TMH4 Gly107-Phe111	62.45	NDUFA9 NADPH	16.59
249	ALA	NDUFA9		3	ND3-Cys39	28.05	ND2 TMH4 Gly107-Phe111	45.73	NDUFA9 NADPH	19.59
342	ILE	NDUFA9		3	ND3-Cys39	51.05	ND2 TMH4 Gly107-Phe111	64.97	NDUFA9 NADPH	23.16
33	ARG	NDUFB1		3	ND4L Glu70	61.37	ND2 TMH9 Gly252-Gly259	28.06	NDUFB10 Cys70&Cys77	22.05
38	ARG	NDUFB1		3	ND4L Glu70	71.26	ND2 TMH9 Gly252-Gly259	37.92	NDUFB10 Cys70&Cys77	16.15
34	VAL	NDUFB10		3	ND4L Glu70	113.60	ND4 TMH12 Leu367-Thr372	31.69	NDUFB10 Cys106&Cys118	30.79
65	PRO	NDUFB10		3	ND4L Glu70	76.85	ND4 TMH7 Met207-Leu214	38.61	NDUFB10 Cys70&Cys77	7.04
109	ARG	NDUFB10		3	ND4L Glu70	106.01	ND4 TMH12 Leu367-Thr372	28.41	NDUFB10 Cys106&Cys118	5.41
139	LEU	NDUFB10		3	ND4L Glu70	72.99	ND4 TMH7 Met207-Leu214	22.93	NDUFB11 Cys112&NDUFB10 Cys148	7.31
164	LYS	NDUFB10		3	ND4L Glu70	89.26	ND4 Glu378	48.81	NDUFB11 Cys112&NDUFB10 Cys148	24.83
55	VAL	NDUFB11		3	ND4L Glu70	80.10	ND4 Ala223 pi bulge	29.83	NDUFA10 ADP	55.64
56	TRP	NDUFB11		3	ND4L Glu70	78.03	ND4 Ala223 pi bulge	29.59	NDUFA10 ADP	55.94
10	ARG	NDUFB2		3	ND4L Glu70	124.68	ND5 Lys336	33.54	NDUFB9 LYK (18-20)	28.15
60	VAL	NDUFB3		3	ND4L Glu70	130.29	ND5 TMH12 Gly382-Tyr390	21.34	NDUFAB1 (β)-acyl chain	27.29
2	PHE	NDUFB4		3	ND4L Glu70	107.07	ND5 His230	31.61	NDUFAB1 (β)-acyl chain	12.74
25	SER	NDUFB4		3	ND4L Glu70	96.39	ND5 Asp554 pi bulge	47.02	NDUFAB1 (β)-acyl chain	19.37
66	ARG	NDUFB4		3	ND4L Glu70	68.86	ND5 Asp554 pi bulge	19.54	NDUFB9 LYK (18-20)	39.62
11	ILE	NDUFB5		3	ND4L Glu70	96.01	ND5 His248	30.40	NDUFB9 LYK (18-20)	33.87
87	TYR	NDUFB5		3	ND4L Glu70	70.59	ND4 Lys237	29.80	NDUFB10 Cys70&Cys77	13.26
137	PRO	NDUFB5		3	ND3 Glu68	34.44	ND2 Glu34	35.17	NDUFS5 Cys32&Cys65	16.27
10	ARG	NDUFB6		3	ND4L Glu70	103.18	ND5 His248	46.61	NDUFB9 LYK (18-20)	23.62
20	ARG	NDUFB6		3	ND4L Glu70	105.96	ND5 His248	44.27	NDUFB9 LYK (18-20)	30.03
25	GLN	NDUFB6		3	ND4L Glu70	108.19	ND5 His248	39.22	NDUFB9 LYK (18-20)	35.09
42	MET	NDUFB6		3	ND4L Glu70	114.43	ND4 TMH12 Leu367-Thr372	43.36	NDUFB7 Cys68&Cys79	58.21

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
80	ILE	NDUFB6		3	ND4L Glu70	104.02	ND4 TMH12 Leu367-Thr372	23.77	NDUFB10 Cys106&Cys118	30.11
108	THR	NDUFB6		3	ND4L Glu70	125.75	ND5 TMH12 Gly382-Tyr390	34.18	NDUFB7 Cys68&Cys79	15.91
29	GLY	NDUFB7		3	ND4L Glu70	129.92	ND5 TMH12 Gly382-Tyr390	40.31	NDUFB7 Cys58-89	25.53
43	GLN	NDUFB7		3	ND4L Glu70	118.85	ND5 TMH12 Gly382-Tyr390	32.52	NDUFB7 Cys58-89	14.69
48	ALA	NDUFB7		3	ND4L Glu70	113.48	ND5 His328	25.17	NDUFB10 Cys106&Cys118	12.43
55	ARG	NDUFB7		3	ND4L Glu70	118.13	ND5 TMH12 Gly382-Tyr390	23.39	NDUFB7 Cys58-89	9.72
61	HIS	NDUFB7		3	ND4L Glu70	129.78	ND5 TMH12 Gly382-Tyr390	21.40	NDUFB7 Cys58-89	5.43
70	ARG	NDUFB7		3	ND4L Glu70	123.92	ND5 TMH12 Gly382-Tyr390	20.00	NDUFB7 Cys68&Cys79	5.40
83	ARG	NDUFB7		3	ND4L Glu70	128.25	ND5 TMH12 Gly382-Tyr390	13.30	NDUFB7 Cys68&Cys79	7.11
92	ARG	NDUFB7		3	ND4L Glu70	131.33	ND5 TMH12 Gly382-Tyr390	18.24	NDUFB7 Cys58-89	5.15
105	ARG	NDUFB7		3	ND4L Glu70	127.87	ND5 TMH12 Gly382-Tyr390	31.49	NDUFB7 Cys58-89	25.55
110	ARG	NDUFB7		3	ND4L Glu70	133.02	ND5 TMH12 Gly382-Tyr390	39.51	NDUFB7 Cys58-89	32.04
112	LYS	NDUFB7		3	ND4L Glu70	130.11	ND5 TMH12 Gly382-Tyr390	42.18	ND5 Cys279&NDUFB8 Cys115	34.28
28	MET	NDUFB8		3	ND4L Glu70	77.85	ND5 Asp554 pi bulge	29.02	NDUFB9 LYK (18-20)	31.15
38	ASP	NDUFB8		3	ND4L Glu70	67.21	ND5 Asp554 pi bulge	18.29	NDUFA11 Cys94&Cys114	27.75
75	PRO	NDUFB8		3	ND4L Glu70	76.90	ND5 Asp554 pi bulge	23.75	NDUFB9 LYK (18-20)	30.87
87	ARG	NDUFB8		3	ND4L Glu70	84.28	ND5 TMH7 Gly222-Pro289	15.43	NDUFB9 LYK (18-20)	28.78
93	THR	NDUFB8		3	ND4L Glu70	96.95	ND5 TMH7 Gly222-Pro288	16.87	NDUFAB1 (β)-acyl chain	15.05
94	PRO	NDUFB8		3	ND4L Glu70	100.15	ND5 TMH7 Gly222-Pro288	19.19	NDUFAB1 (β)-acyl chain	12.16
3	LEU	NDUFB9		3	ND4L Glu70	101.46	ND5 TMH7 Gly222-Pro288	47.49	NDUFAB1 (β)-acyl chain	13.21
7	PRO	NDUFB9		3	ND4L Glu70	94.61	ND5 TMH7 Gly222-Pro288	39.76	NDUFB9 LYK (18-20)	17.56
36	TYR	NDUFB9		3	ND4L Glu70	119.85	ND5 TMH7 Gly222-Pro288	32.50	NDUFAB1 (β)-acyl chain	9.99
46	ARG	NDUFB9		3	ND4L Glu70	118.88	ND5 TMH7 Gly222-Pro288	40.35	NDUFAB1 (β)-acyl chain	5.36
128	ARG	NDUFB9		3	ND4L Glu70	93.57	ND4 Ala223 pi bulge	42.34	NDUFB9 LYK (18-20)	36.93
143	THR	NDUFB9		3	ND4L Glu70	92.65	ND4 Ala223 pi bulge	49.50	NDUFAB1 (β)-acyl chain	31.65
174	ARG	NDUFB9		3	ND4L Glu70	99.94	ND5 His248	30.38	NDUFB9 LYK (18-20)	24.04
176	ARG	NDUFB9		3	ND4L Glu70	103.08	ND5 His248	26.93	NDUFB9 LYK (18-20)	22.36
178	MET	NDUFB9		3	ND4L Glu70	100.49	ND5 His248	25.87	NDUFB9 LYK (18-20)	27.96
41	GLU	NDUFC1		3	ND4L Glu70	53.26	ND2 TMH9 Gly252-Gly259	35.42	NDUFB10 Cys70&Cys77	40.95

Residue number	Identity	Subunit	Mutation class	Polymorphism class	Group A features: Core nuclear & ND1, 3, 4L & 6 subunits	Distance (Å)	Group B features: ND2, 4 & 5 subunits	Distance (Å)	Group C features: Supernumerary subunits	Distance (Å)
35	PHE	NDUFC2		3	ND4L Glu70	47.72	ND2 TMH9 Gly252-Gly259	14.20	NDUFB10 Cys70&Cys77	41.47
59	ARG	NDUFC2		3	ND4L Glu70	46.71	ND2 TMH9 Gly252-Gly259	22.92	NDUFA10 ADP	45.98
119	ARG	NDUFC2		3	ND4L Glu70	74.11	ND4 Glu378	20.01	NDUFB11 Cys112&NDUFB10 Cys148	14.40
6	SER	NDUFS1		3	Fe-S cluster	27.71	ND2 Leu131 pi bulge	124.07	NDUFS6 Zinc	62.54
122	MET	NDUFS1		3	Fe-S cluster	13.50	ND2 Glu34	103.41	NDUFS6 Zinc	30.73
280	THR	NDUFS1		3	Fe-S cluster	34.13	ND2 TMH4 Gly107-Phe111	99.77	NDUFA9 NADPH	37.67
331	LEU	NDUFS1		3	Fe-S cluster	48.37	ND2 TMH4 Gly107-Phe111	104.67	NDUFAB1 (α)-acyl chain	43.15
334	LEU	NDUFS1		3	Fe-S cluster	52.97	ND2 TMH4 Gly107-Phe111	102.81	NDUFAB1 (α)-acyl chain	42.07
479	MET	NDUFS1		3	Fe-S cluster	49.47	ND2 His112	138.27	NDUFAB1 (α)-acyl chain	71.38
515	ARG	NDUFS1		3	Fe-S cluster	53.71	ND2 TMH4 Gly107-Phe111	123.90	NDUFAB1 (α)-acyl chain	63.85
634	ASP	NDUFS1		3	Fe-S cluster	31.97	ND2 His112	108.88	NDUFAB1 (α)-acyl chain	43.59
646	ASN	NDUFS1		3	Fe-S cluster	49.47	ND2 His112	130.88	NDUFAB1 (α)-acyl chain	64.46
689	LYS	NDUFS1		3	Fe-S cluster	25.35	ND2 Glu34	123.28	NDUFS6 Zinc	33.45
80	ILE	NDUFS2		3	Q-site	10.35	ND2 Glu34	58.14	NDUFAB1 (α)-acyl chain	22.70
379	VAL	NDUFS2		3	Fe-S cluster	24.00	ND2 Glu34	82.71	NDUFAB1 (α)-acyl chain	36.57
409	HIS	NDUFS2		3	NDUFS2 D160	15.77	ND2 Glu34	41.42	NDUFAB1 (α)-acyl chain	35.96
100	ARG	NDUFS3		3	NDUFS2 Y108	32.82	ND2 Glu34	74.09	NDUFA6 LYR (34-36)	48.74
94	SER	NDUFS4		3	Fe-S cluster	39.93	ND2 His112	87.82	NDUFAB1 (α)-acyl chain	24.68
115	SER	NDUFS4		3	Fe-S cluster	15.94	ND2 Glu34	123.13	NDUFS6 Zinc	48.66
71	THR	NDUFS5		3	ND4L Glu34	36.77	ND2 Glu34	36.32	NDUFS5 Cys32&Cys65	9.70
92	THR	NDUFS5		3	ND4L Glu34	52.69	ND2 Glu34	56.26	NDUFS5 Cys32&Cys65	32.70
97	HIS	NDUFS5		3	ND4L Glu34	47.24	ND2 Glu34	50.28	NDUFA8 Cys45&Cys55	26.75
133	ASP	NDUFS7		3	Fe-S cluster	14.93	ND2 Glu34	69.12	NDUFA9 NADPH	23.23
134	ARG	NDUFS7		3	Fe-S cluster	16.73	ND2 Glu34	68.65	NDUFA9 NADPH	23.09
148	THR	NDUFS7		3	Fe-S cluster	10.20	ND2 Glu34	72.00	NDUFA9 NADPH	38.42
11	SER	NDUFV1		3	FMN	26.76	ND2 Glu34	154.51	NDUFS6 Zinc	69.07
32	ARG	NDUFV1		3	FMN	23.45	ND2 Glu34	167.69	NDUFS6 Zinc	67.16
168	GLY	NDUFV1		3	FMN	27.80	ND2 Glu34	169.25	NDUFS6 Zinc	75.73
175	VAL	NDUFV1		3	FMN	18.13	ND2 Glu34	157.00	NDUFS6 Zinc	60.14
339	ARG	NDUFV1		3	Fe-S cluster	18.30	ND2 Glu34	140.65	NDUFS6 Zinc	48.26
376	MET	NDUFV1		3	Fe-S cluster	17.96	ND2 Glu34	125.58	NDUFS6 Zinc	29.41
385	ARG	NDUFV1		3	Fe-S cluster	27.88	ND2 Glu34	117.39	NDUFS6 Zinc	30.44
418	LEU	NDUFV1		3	Fe-S cluster	19.54	ND2 Glu34	129.65	NDUFS6 Zinc	40.50
19	THR	NDUFV2		3	Fe-S cluster	31.54	ND2 Glu34	146.18	NDUFS6 Zinc	40.59
76	PRO	NDUFV2		3	Fe-S cluster	19.12	ND2 Glu34	137.45	NDUFS6 Zinc	40.03
140	ILE	NDUFV2		3	Fe-S cluster	12.99	ND2 Glu34	142.83	NDUFS6 Zinc	35.60

Appendix 7.9D Type of residue change for complex I subunit gene variants (with variant class reclassification). Residue number refers to the mature protein.

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type	Polymorphism class	Residue change	Polymorphism change type
340	ARG	ND4	1	R-H	Mild	2	R-H	Mild
64	MET	ND6	1	M-I	Mild	2	M-V	Mild
52	ALA	ND1	1	A-T	Mild	3	A-T	Mild
110	SER	ND1	1	S-N	Medium	3	S-N	Medium
132	ALA	ND1	1	A-T	Mild	3	A-T	Mild
289	LEU	ND1	1	L-M	Mild	3	L-Q	Severe
34	SER	ND3	1	S-P	Medium	3	S-F	Severe
45	SER	ND3	1	S-P	Medium	3	S-Y	Severe
47	ALA	ND3	1	A-T	Mild	3	A-T	Mild
36	GLY	ND6	1	G-S	Medium	3	G-S	Medium
72	ALA	ND6	1	A-V	Medium	3	A-P	Medium
259	MET	NDUFS2	1	M-T	Severe	3	M-T	Severe
393	ALA	NDUFS2	1	A-S	Mild	3	A-S	Mild
30	TYR	ND1	2	Y-H	Mild	1	Y-H	Mild
1	MET	ND1	2	M-K	Severe	2	M-T	Severe
31	MET	ND1	2	M-V	Mild	2	M-V	Mild
164	THR	ND1	2	T-A	Mild	2	T-A	Mild
240	THR	ND1	2	T-M	Severe	2	T-A	Mild
57	ILE	ND2	2	I-M	Mild	2	I-M	Mild
60	ILE	ND3	2	I-T	Severe	2	I-T	Severe
109	THR	ND4	2	T-A	Mild	2	T-A	Mild
165	ILE	ND4	2	I-T	Severe	2	I-T	Severe
313	VAL	ND4	2	V-I	Mild	2	V-I	Mild
1	MET	ND5	2	M-T	Severe	2	M-T	Severe
159	TYR	ND5	2	Y-H	Mild	2	Y-H	Mild
434	GLN	ND5	2	Q-R	Mild	2	Q-R	Mild
132	SER	ND6	2	S-A	Mild	2	S-P	Medium
290	ARG	NDUFS2	2	R-Q	Mild	2	R-Q	Mild
28	LEU	ND1	2	L-M	Mild	3	L-M	Mild
277	TYR	ND1	2	Y-C	Severe	3	Y-H	Mild
409	TYR	ND4	2	Y-H	Mild	3	Y-H	Mild
26	ILE	ND6	2	I-M	Mild	3	I-V	Mild
117	ASN	ND6	2	N-D	Mild	3	N-D	Mild
21	TRP	NDUFB3	2	L-P	Severe	3	W-R	Severe
179	ARG	NDUFV1	2	R-C	Severe	3	R-C	Severe
366	ARG	NDUFV1	2	R-C	Severe	3	R-C	Severe
4	ALA	ND1	3	A-T	Mild	1	A-T	Mild
304	TYR	ND1	3	Y-H	Mild	1	Y-H	Mild
150	ASN	ND2	3	N-D	Mild	1	N-D	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type	Polymorphism class	Residue change	Polymorphism change type
331	ALA	ND2	3	A-T	Mild	1	A-T	Mild
458	ALA	ND5	3	A-T	Mild	1	A-T	Mild
544	THR	ND5	3	T-M	Severe	1	T-A	Mild
119	ASN	ND6	3	N-D	Mild	1	N-S	Medium
11	VAL	ND1	3	V-A	Medium	2	V-A	Medium
31	MET	ND1	3	M-I	Mild	2	M-T	Severe
39	VAL	ND1	3	V-I	Mild	2	V-I	Mild
64	ALA	ND1	3	A-S	Mild	2	A-S	Mild
64	ALA	ND1	3	A-V	Medium	2	A-V	Medium
113	VAL	ND1	3	V-A	Medium	2	V-A	Medium
144	VAL	ND1	3	V-I	Mild	2	V-I	Mild
147	ALA	ND1	3	A-L	Severe	2	A-T	Mild
64	ALA	ND2	3	A-T	Mild	2	A-T	Mild
122	THR	ND2	3	T-A	Mild	2	T-A	Mild
10	ASN	ND3	3	N-D	Mild	2	N-D	Mild
423	ILE	ND4	3	I-V	Mild	2	I-V	Mild
9	THR	ND5	3	T-A	Mild	2	T-A	Mild
21	THR	ND5	3	T-A	Mild	2	T-A	Mild
100	ILE	ND5	3	I-V	Mild	2	I-V	Mild
267	ALA	ND5	3	A-T	Mild	2	A-T	Mild
398	THR	ND5	3	T-A	Mild	2	T-A	Mild
33	ILE	ND6	3	I-V	Mild	2	I-V	Mild
58	ILE	ND6	3	I-T	Severe	2	I-V	Mild
32	GLY	NDUFA1	3	G-R	Severe	2	G-R	Severe
103	ARG	NDUFA11	3	R-L	Severe	2	R-L	Severe
408	LEU	NDUFS1	3	L-V	Mild	2	L-V	Mild
2	PRO	ND1	3	P-L	Severe	3	P-L	Severe
30	TYR	ND1	3	Y-C	Severe	3	Y-C	Severe
31	MET	ND1	3	M-T	Severe	3	M-I	Mild
128	ALA	ND1	3	A-T	Mild	3	A-P	Medium
176	LEU	ND1	3	L-M	Mild	3	L-M	Mild
187	ILE	ND1	3	I-T	Severe	3	I-V	Mild
230	ASN	ND1	3	N-S	Medium	3	N-S	Medium
276	ALA	ND1	3	A-T	Mild	3	A-T	Mild
328	THR	ND2	3	T-M	Severe	3	T-M	Severe
331	ALA	ND2	3	A-S	Mild	3	A-V	Medium
71	ALA	ND4L	3	A-T	Mild	3	A-T	Mild
99	SER	ND5	3	S-T	Mild	3	S-A	Mild
149	ILE	ND5	3	I-S	Severe	3	I-V	Mild
576	ILE	ND5	3	I-T	Severe	3	I-T	Severe
33	ILE	ND6	3	I-V	Mild	3	I-L	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type	Polymorphism class	Residue change	Polymorphism change type
112	VAL	ND6	3	V-M	Mild	3	V-M	Mild
171	ALA	ND6	3	A-T	Mild	3	A-V	Medium
100	LEU	NDUFA10	3	L-S	Severe	3	L-S	Severe
251	GLN	NDUFA10	3	Q-E	Mild	3	Q-E	Mild
40	ARG	NDUFA9	3	R-L	Severe	3	R-L	Severe
40	ARG	NDUFA9	3	R-H	Mild	3	R-H	Mild
50	ARG	NDUFA9	3	R-W	Severe	3	R-W	Severe
279	ILE	NDUFA9	3	I-M	Mild	3	I-M	Mild
157	ARG	NDUFB9	3	R-G	Severe	3	R-G	Severe
154	ILE	NDUFS1	3	I-M	Mild	3	I-V	Mild
181	MET	NDUFS1	3	M-I	Mild	3	M-T	Severe
483	VAL	NDUFS1	3	V-I	Mild	3	V-I	Mild
511	VAL	NDUFS1	3	V-M	Mild	3	V-M	Mild
287	VAL	NDUFS2	3	V-A	Medium	3	V-A	Medium
347	HIS	NDUFS2	3	H-D	Medium	3	H-D	Medium
210	ARG	NDUFS3	3	R-C	Severe	3	R-C	Severe
58	ALA	NDUFS6	3	A-V	Medium	3	A-V	Medium
60	ASP	NDUFS6	3	D-N	Medium	3	D-N	Mild
133	ALA	NDUFV1	3	A-T	Mild	3	A-T	Mild
337	MET	NDUFV1	3	M-L	Mild	3	M-L	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
24	GLU	ND1	1	E-K	Mild
131	GLY	ND1	1	G-S	Medium
143	GLU	ND1	1	E-K	Mild
195	ARG	ND1	1	R-Q	Mild
340	ARG	ND4	1	R-S	Severe
65	VAL	ND4L	1	V-A	Medium
124	PHE	ND5	1	F-L	Mild
393	ASP	ND5	1	D-G	Medium
393	ASP	ND5	1	D-N	Medium
60	LEU	ND6	1	L-S	Severe
63	MET	ND6	1	M-V	Mild
64	MET	ND6	1	M-V	Mild
8	GLY	NDUFA1	1	G-R	Severe
37	ARG	NDUFA1	1	R-S	Severe
107	GLN	NDUFA10	1	Q-R	Mild
286	ARG	NDUFA9	1	R-P	Severe
208	LEU	NDUFS1	1	L-V	Mild
218	ARG	NDUFS1	1	R-W	Severe
229	ASP	NDUFS1	1	D-G	Medium
385	ARG	NDUFS1	1	R-C	Severe
572	THR	NDUFS1	1	T-A	Mild
596	ASP	NDUFS1	1	D-N	Medium
195	ARG	NDUFS2	1	R-Q	Mild
196	PRO	NDUFS2	1	P-Q	Severe
380	SER	NDUFS2	1	S-P	Medium
109	THR	NDUFS3	1	T-I	Severe
163	ARG	NDUFS3	1	R-W	Severe
87	CYS	NDUFS6	1	C-Y	Severe
84	VAL	NDUFS7	1	V-M	Mild
29	GLU	NDUFS8	1	E-Q	Mild
43	ARG	NDUFS8	1	R-W	Severe
45	PRO	NDUFS8	1	P-L	Severe
51	PRO	NDUFS8	1	P-L	Severe
68	ARG	NDUFS8	1	R-H	Mild
104	ARG	NDUFS8	1	R-H	Mild
125	ALA	NDUFS8	1	A-D	Severe
97	ALA	NDUFV1	1	A-T	Mild
102	PRO	NDUFV1	1	P-L	Severe
186	CYS	NDUFV1	1	C-G	Severe
194	GLU	NDUFV1	1	E-K	Mild
403	THR	NDUFV1	1	T-M	Severe

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
59	GLU	ND1	2	E-K	Mild
208	VAL	ND1	2	V-L	Mild
214	GLU	ND1	2	E-K	Mild
215	TYR	ND1	2	Y-H	Mild
285	LEU	ND1	2	L-P	Severe
71	LEU	ND2	2	L-M	Mild
259	GLY	ND2	2	G-S	Medium
26	GLN	ND3	2	Q-K	Mild
66	ASP	ND3	2	D-N	Medium
32	CYS	ND4L	2	C-R	Severe
145	GLU	ND5	2	E-G	Severe
171	ALA	ND5	2	A-V	Medium
236	ALA	ND5	2	A-T	Mild
237	MET	ND5	2	M-L	Mild
250	SER	ND5	2	S-C	Severe
465	GLY	ND5	2	G-E	Severe
59	TYR	ND6	2	Y-C	Severe
74	ALA	ND6	2	A-V	Medium
19	PRO	NDUFA1	2	P-S	Medium
64	GLY	NDUFA10	2	G-E	Severe
259	LEU	NDUFA10	2	L-P	Severe
56	ARG	NDUFA13	2	R-H	Mild
44	LYS	NDUFA2	2	K-T	Medium
108	GLU	NDUFA8	2	E-K	Mild
92	GLU	NDUFB11	2	E-K	Mild
63	LEU	NDUFB9	2	Y-C	Severe
96	TYR	NDUFB9	2	V-D	Severe
48	VAL	NDUFS1	2	G-E	Severe
143	GLY	NDUFS1	2	V-A	Medium
205	VAL	NDUFS1	2	Q-K	Mild
499	GLN	NDUFS1	2	Y-C	Severe
672	TYR	NDUFS1	2	M-V	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
684	MET	NDUFS1	2	Y-C	Severe
77	ASP	NDUFS2	2	D-V	Severe
85	ARG	NDUFS2	2	R-Q	Mild
105	ARG	NDUFS2	2	R-Q	Mild
115	GLU	NDUFS2	2	E-K	Mild
191	ALA	NDUFS2	2	A-V	Medium
300	ARG	NDUFS2	2	R-Q	Mild
410	MET	NDUFS2	2	M-K	Severe
413	ASP	NDUFS2	2	D-N	Medium
187	PRO	NDUFS3	2	P-L	Severe
72	TRP	NDUFS4	2	W-R	Severe
77	ASP	NDUFS4	2	D-H	Medium
95	PRO	NDUFS5	2	P-S	Medium
67	ARG	NDUFS7	2	R-C	Severe
107	ARG	NDUFS7	2	R-H	Mild
20	ARG	NDUFS8	2	R-W	Severe
60	ARG	NDUFS8	2	R-C	Severe
120	GLY	NDUFS8	2	G-S	Medium
36	SER	NDUFV1	2	S-P	Medium
68	ARG	NDUFV1	2	R-G	Severe
91	LYS	NDUFV1	2	K-E	Mild
127	ARG	NDUFV1	2	R-W	Severe
184	TYR	NDUFV1	2	Y-C	Severe
191	ALA	NDUFV1	2	A-V	Medium
226	GLU	NDUFV1	2	E-K	Mild
232	PRO	NDUFV1	2	P-L	Severe
321	ALA	NDUFV1	2	A-V	Medium
357	GLU	NDUFV1	2	E-K	Mild
358	SER	NDUFV1	2	S-R	Severe
412	ALA	NDUFV1	2	A-P	Medium
177	LYS	NDUFV2	2	K-R	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
12	PRO	ND1	3	P-S	Medium
25	ARG	ND1	3	R-Q	Mild
34	ARG	ND1	3	R-H	Mild
38	ASN	ND1	3	N-D	Mild
56	PHE	ND1	3	F-L	Mild
143	GLU	ND1	3	E-Q	Mild
218	GLY	ND1	3	G-D	Medium
279	ARG	ND1	3	R-Q	Mild
55	ALA	ND2	3	A-G	Medium
60	PHE	ND2	3	F-S	Severe
128	LEU	ND2	3	L-Q	Severe
158	LEU	ND4	3	L-P	Severe
206	LYS	ND4	3	K-Q	Mild
372	THR	ND4	3	T-N	Medium
387	SER	ND4	3	S-F	Severe
412	THR	ND4	3	T-I	Severe
423	ILE	ND4	3	I-T	Severe
25	HIS	ND4L	3	H-R	Mild
41	PHE	ND4L	3	F-C	Severe
96	VAL	ND5	3	V-I	Mild
239	GLY	ND5	3	G-S	Medium
243	VAL	ND5	3	V-I	Mild
253	VAL	ND5	3	V-A	Medium
312	LEU	ND5	3	L-P	Severe
348	HIS	ND5	3	N-S	Medium
392	LYS	ND5	3	K-M	Severe
415	ALA	ND5	3	A-G	Medium
499	LEU	ND5	3	L-M	Mild
505	ASN	ND5	3	N-H	Mild
585	LYS	ND5	3	K-N	Mild
25	PRO	ND6	3	P-L	Severe
26	ILE	ND6	3	I-T	Severe
79	PRO	ND6	3	P-S	Medium
82	TRP	ND6	3	W-R	Severe
101	GLY	ND6	3	G-V	Severe
43	TYR	NDUFA1	3	Y-H	Mild
173	ASP	NDUFA10	3	D-N	Medium
176	VAL	NDUFA10	3	V-I	Mild

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
9	TRP	NDUFA11	3	W-C	Severe
77	ALA	NDUFA11	3	A-T	Mild
81	GLU	NDUFA11	3	E-G	Severe
68	TYR	NDUFA2	3	Y-C	Severe
26	ALA	NDUFA9	3	A-V	Medium
33	TYR	NDUFA9	3	Y-C	Severe
287	MET	NDUFA9	3	M-T	Severe
303	LEU	NDUFA9	3	L-F	Mild
60	THR	NDUFB9	3	T-S	Mild
39	ARG	NDUFS1	3	R-Q	Mild
230	VAL	NDUFS1	3	V-G	Severe
286	ASN	NDUFS1	3	N-S	Medium
432	ILE	NDUFS1	3	I-V	Mild
452	VAL	NDUFS1	3	V-F	Medium
543	ILE	NDUFS1	3	I-F	Mild
553	GLY	NDUFS1	3	G-E	Severe
577	GLU	NDUFS1	3	E-D	Mild
680	ALA	NDUFS1	3	A-T	Mild
20	TYR	NDUFS2	3	Y-C	Severe
376	VAL	NDUFS2	3	V-M	Mild
42	ARG	NDUFS4	3	R-H	Mild
49	VAL	NDUFS6	3	V-M	Mild
55	ARG	NDUFS6	3	R-W	Severe
74	ASN	NDUFS6	3	N-I	Severe
77	ARG	NDUFS7	3	R-C	Severe
81	VAL	NDUFS7	3	V-I	Mild
167	ARG	NDUFS7	3	R-G	Severe
71	PRO	NDUFS8	3	P-S	Medium
81	LYS	NDUFS8	3	K-E	Mild
109	TYR	NDUFS8	3	Y-F	Mild
127	PRO	NDUFS8	3	P-T	Medium
170	GLN	NDUFS8	3	Q-L	Severe
19	ASP	NDUFV1	3	D-G	Medium
75	THR	NDUFV1	3	T-A	Mild
131	ALA	NDUFV1	3	A-T	Mild
140	GLY	NDUFV1	3	G-E	Severe
247	ARG	NDUFV1	3	R-K	Mild
263	ASN	NDUFV1	3	N-S	Medium

Residue number	Identity	Subunit	Mutation class	Residue change	Mutation change type
297	VAL	NDUFV1	3	V-A	Medium
348	ALA	NDUFV1	3	A-T	Mild
153	MET	NDUFV2	3	M-V	Mild
158	ASP	NDUFV2	3	D-G	Medium

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
67	THR	ND1	1	T-A	Mild
81	ILE	ND1	1	I-V	Mild
309	ILE	ND1	1	I-T	Severe
8	VAL	ND2	1	V-I	Mild
119	THR	ND2	1	T-A	Mild
193	VAL	ND2	1	V-I	Mild
237	LEU	ND2	1	L-M	Mild
325	PHE	ND2	1	F-L	Mild
88	VAL	ND3	1	V-A	Medium
114	THR	ND3	1	T-A	Mild
140	PRO	ND4	1	P-S	Medium
404	ALA	ND4	1	A-T	Mild
47	MET	ND4L	1	M-T	Severe
8	THR	ND5	1	T-A	Mild
24	VAL	ND5	1	V-I	Mild
257	ILE	ND5	1	I-V	Mild
314	MET	ND5	1	M-V	Mild
475	ALA	ND5	1	A-T	Mild
482	ILE	ND5	1	I-V	Mild
485	TYR	ND5	1	Y-H	Mild
531	SER	ND5	1	S-T	Mild
533	THR	ND5	1	T-M	Severe
555	LEU	ND5	1	L-Q	Severe
166	ILE	ND6	1	I-V	Mild
47	LYS	NDUFA7	1	K-Q	Mild
259	PRO	NDUFA9	1	P-L	Severe
145	PRO	NDUFB9	1	P-S	Medium
46	LEU	NDUFC2	1	L-V	Mild
102	VAL	NDUFV2	1	V-A	Medium

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
11	VAL	ND1	2	V-M	Mild
29	GLY	ND1	2	G-S	Medium
43	TYR	ND1	2	Y-C	Severe
64	ALA	ND1	2	A-T	Mild
69	THR	ND1	2	T-A	Mild
87	THR	ND1	2	T-A	Mild
89	LEU	ND1	2	L-F	Mild
164	THR	ND1	2	T-S	Mild
187	ILE	ND1	2	I-T	Severe
229	THR	ND1	2	T-M	Severe
240	THR	ND1	2	T-M	Severe
248	ASP	ND1	2	D-N	Mild
263	THR	ND1	2	T-A	Mild
275	THR	ND1	2	T-A	Mild
277	TYR	ND1	2	Y-C	Severe
307	MET	ND1	2	M-V	Mild
13	ILE	ND2	2	I-V	Mild
31	VAL	ND2	2	V-A	Medium
43	VAL	ND2	2	V-I	Mild
57	ILE	ND2	2	I-T	Severe
88	ASN	ND2	2	N-S	Medium
100	MET	ND2	2	M-V	Mild
152	SER	ND2	2	S-N	Medium
164	ALA	ND2	2	A-T	Mild
164	ALA	ND2	2	A-V	Medium
202	ILE	ND2	2	I-T	Severe
203	LEU	ND2	2	L-F	Mild
265	ALA	ND2	2	A-T	Mild
265	ALA	ND2	2	A-V	Medium
270	PHE	ND2	2	F-L	Mild
278	ILE	ND2	2	I-T	Severe
278	ILE	ND2	2	I-V	Mild
284	THR	ND2	2	T-S	Mild
288	LEU	ND2	2	L-I	Mild
342	PHE	ND2	2	F-L	Mild
9	ILE	ND3	2	I-T	Severe
29	GLY	ND3	2	G-S	Medium
45	SER	ND3	2	S-F	Severe
49	VAL	ND3	2	V-I	Mild
88	VAL	ND3	2	V-I	Mild
96	ILE	ND3	2	I-T	Severe
103	ALA	ND3	2	A-T	Mild
47	ASN	ND4	2	N-S	Medium
50	PHE	ND4	2	F-L	Mild
54	PRO	ND4	2	P-L	Severe
86	SER	ND4	2	S-N	Medium
89	LEU	ND4	2	L-P	Severe
101	SER	ND4	2	S-F	Severe
110	PHE	ND4	2	F-L	Mild
131	ALA	ND4	2	A-T	Mild
131	ALA	ND4	2	A-V	Medium
138	ASN	ND4	2	N-S	Medium
149	PHE	ND4	2	F-L	Mild
230	VAL	ND4	2	V-M	Mild
299	THR	ND4	2	T-A	Mild
390	ASN	ND4	2	N-S	Medium
402	VAL	ND4	2	V-I	Mild
424	ASN	ND4	2	N-S	Medium
442	SER	ND4	2	S-F	Severe
442	SER	ND4	2	S-A	Mild
445	LEU	ND4	2	L-F	Mild
459	SER	ND4	2	S-Y	Severe
13	THR	ND4L	2	T-A	Mild
94	ASN	ND4L	2	N-S	Medium
4	HIS	ND5	2	H-Y	Mild
9	THR	ND5	2	T-I	Severe
13	THR	ND5	2	T-A	Mild
23	LEU	ND5	2	L-F	Mild
69	ALA	ND5	2	A-V	Medium
74	THR	ND5	2	T-I	Severe
182	PHE	ND5	2	F-L	Mild
201	MET	ND5	2	M-V	Mild
202	ALA	ND5	2	A-T	Mild
205	ASN	ND5	2	N-S	Medium
205	ASN	ND5	2	N-T	Medium
261	ILE	ND5	2	I-V	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
265	PRO	ND5	2	P-S	Medium
270	SER	ND5	2	S-N	Medium
283	ILE	ND5	2	I-V	Mild
410	SER	ND5	2	S-F	Severe
432	THR	ND5	2	T-A	Mild
439	THR	ND5	2	T-A	Mild
449	THR	ND5	2	T-A	Mild
459	ALA	ND5	2	A-T	Mild
469	THR	ND5	2	T-A	Mild
471	ASN	ND5	2	N-S	Medium
478	PHE	ND5	2	F-L	Mild
495	PHE	ND5	2	F-L	Mild
515	SER	ND5	2	S-Y	Severe
515	SER	ND5	2	S-P	Medium
518	CYS	ND5	2	C-Y	Severe
530	PRO	ND5	2	P-S	Medium
531	SER	ND5	2	S-N	Medium
533	THR	ND5	2	T-A	Mild
536	THR	ND5	2	T-A	Mild
538	PRO	ND5	2	P-S	Medium
541	GLY	ND5	2	G-A	Medium
544	THR	ND5	2	T-M	Severe
549	PRO	ND5	2	P-S	Medium
556	THR	ND5	2	T-A	Mild
573	THR	ND5	2	T-A	Mild
575	ILE	ND5	2	I-V	Mild
592	PHE	ND5	2	F-L	Mild
598	THR	ND5	2	T-I	Severe
598	THR	ND5	2	T-A	Mild
14	MET	ND6	2	M-V	Mild
31	VAL	ND6	2	V-A	Medium
86	VAL	ND6	2	V-A	Medium
97	ALA	ND6	2	A-V	Medium
119	ASN	ND6	2	N-D	Mild
120	SER	ND6	2	S-N	Medium
159	THR	ND6	2	T-M	Severe
165	TYR	ND6	2	Y-C	Severe

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
30	ASN	NDUFA10	2	N-S	Medium
13	LEU	NDUFA2	2	L-V	Mild
20	VAL	NDUFA3	2	V-L	Mild
107	HIS	NDUFA6	2	H-Y	Mild
65	PRO	NDUFA7	2	P-A	Medium
18	MET	NDUFB1	2	M-L	Mild
141	ALA	NDUFB10	2	A-T	Mild
52	PRO	NDUFB2	2	P-L	Severe
62	LEU	NDUFB7	2	L-V	Mild
96	SER	NDUFB8	2	S-T	Mild
120	VAL	NDUFB8	2	V-M	Mild
123	VAL	NDUFS3	2	V-L	Mild
8	VAL	NDUFV2	2	V-M	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
2	PRO	ND1	3	P-S	Medium
5	ASN	ND1	3	N-D	Mild
10	ILE	ND1	3	I-T	Severe
10	ILE	ND1	3	I-V	Mild
15	ILE	ND1	3	I-T	Severe
15	ILE	ND1	3	I-V	Mild
17	MET	ND1	3	M-V	Mild
21	MET	ND1	3	M-T	Severe
27	ILE	ND1	3	I-V	Mild
29	GLY	ND1	3	G-A	Medium
68	ILE	ND1	3	I-T	Severe
68	ILE	ND1	3	I-M	Mild
73	THR	ND1	3	T-A	Mild
73	THR	ND1	3	T-S	Mild
76	THR	ND1	3	T-I	Severe
77	LEU	ND1	3	L-M	Mild
81	ILE	ND1	3	I-T	Severe
96	VAL	ND1	3	V-A	Medium
112	ALA	ND1	3	A-T	Mild
116	ILE	ND1	3	I-V	Mild
139	THR	ND1	3	T-A	Mild
147	ALA	ND1	3	A-L	Severe
152	SER	ND1	3	S-A	Mild
153	THR	ND1	3	T-M	Severe
168	THR	ND1	3	T-A	Mild
171	HIS	ND1	3	H-Y	Mild
196	THR	ND1	3	T-A	Mild
213	ILE	ND1	3	I-V	Mild
236	THR	ND1	3	T-I	Severe
236	THR	ND1	3	T-A	Mild
239	THR	ND1	3	T-A	Mild
241	ILE	ND1	3	I-M	Mild
249	ALA	ND1	3	A-V	Medium
258	TYR	ND1	3	Y-C	Severe
260	VAL	ND1	3	V-I	Mild
263	THR	ND1	3	T-I	Severe
265	LEU	ND1	3	L-F	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
266	LEU	ND1	3	L-M	Mild
268	SER	ND1	3	S-F	Severe
273	ILE	ND1	3	I-V	Mild
292	ASN	ND1	3	N-D	Mild
292	ASN	ND1	3	N-S	Medium
300	LEU	ND1	3	L-S	Severe
305	VAL	ND1	3	V-I	Mild
307	MET	ND1	3	M-T	Severe
309	ILE	ND1	3	I-V	Mild
310	THR	ND1	3	T-S	Mild
313	SER	ND1	3	S-G	Medium
318	THR	ND1	3	T-I	Severe
11	SER	ND2	3	S-P	Medium
15	ALA	ND2	3	A-T	Mild
19	ILE	ND2	3	I-T	Severe
31	VAL	ND2	3	V-M	Mild
48	MET	ND2	3	M-T	Severe
49	ASN	ND2	3	N-S	Medium
62	THR	ND2	3	T-M	Severe
69	ILE	ND2	3	I-V	Mild
76	PHE	ND2	3	F-S	Severe
76	PHE	ND2	3	F-L	Mild
78	ASN	ND2	3	N-S	Medium
79	MET	ND2	3	M-T	Severe
80	LEU	ND2	3	L-F	Mild
81	SER	ND2	3	S-F	Severe
85	THR	ND2	3	T-A	Mild
89	THR	ND2	3	T-A	Mild
89	THR	ND2	3	T-N	Medium
94	SER	ND2	3	S-P	Medium
98	ILE	ND2	3	I-M	Mild
100	MET	ND2	3	M-T	Severe
115	VAL	ND2	3	V-I	Mild
125	THR	ND2	3	T-M	Severe
139	ILE	ND2	3	I-L	Mild
141	ILE	ND2	3	I-V	Mild
145	ILE	ND2	3	I-L	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
145	ILE	ND2	3	I-M	Mild
147	PRO	ND2	3	P-S	Medium
148	SER	ND2	3	S-A	Mild
152	SER	ND2	3	S-T	Mild
152	SER	ND2	3	S-G	Medium
156	THR	ND2	3	T-I	Severe
159	ILE	ND2	3	I-V	Mild
166	SER	ND2	3	S-G	Medium
187	MET	ND2	3	M-T	Severe
187	MET	ND2	3	M-V	Mild
200	MET	ND2	3	M-I	Mild
200	MET	ND2	3	M-V	Mild
209	ILE	ND2	3	I-T	Severe
220	ASN	ND2	3	N-D	Mild
220	ASN	ND2	3	N-S	Medium
222	ASN	ND2	3	N-S	Medium
238	THR	ND2	3	T-A	Mild
239	TRP	ND2	3	W-C	Severe
241	THR	ND2	3	T-M	Severe
242	PRO	ND2	3	P-L	Severe
246	SER	ND2	3	S-F	Severe
246	SER	ND2	3	S-P	Medium
266	ILE	ND2	3	I-V	Mild
267	ILE	ND2	3	I-V	Mild
276	LEU	ND2	3	L-F	Mild
277	ILE	ND2	3	I-V	Mild
284	THR	ND2	3	T-A	Mild
285	ILE	ND2	3	I-L	Mild
286	THR	ND2	3	T-S	Mild
317	PHE	ND2	3	F-L	Mild
320	THR	ND2	3	T-A	Mild
323	THR	ND2	3	T-I	Severe
323	THR	ND2	3	T-A	Mild
330	ILE	ND2	3	I-V	Mild
332	LEU	ND2	3	L-F	Mild
333	THR	ND2	3	T-A	Mild
343	MET	ND2	3	M-V	Mild
346	ILE	ND2	3	I-V	Mild
14	ALA	ND3	3	A-S	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
19	ILE	ND3	3	I-T	Severe
24	LEU	ND3	3	L-M	Mild
44	MET	ND3	3	M-V	Mild
89	MET	ND3	3	M-T	Severe
90	SER	ND3	3	S-L	Severe
90	SER	ND3	3	S-T	Mild
90	SER	ND3	3	S-P	Medium
93	LEU	ND3	3	L-S	Severe
97	ILE	ND3	3	I-T	Severe
103	ALA	ND3	3	A-V	Medium
114	THR	ND3	3	T-S	Mild
6	VAL	ND4	3	V-A	Medium
21	HIS	ND4	3	H-Q	Mild
29	THR	ND4	3	T-I	Severe
45	ILE	ND4	3	I-F	Mild
46	ASN	ND4	3	N-D	Mild
50	PHE	ND4	3	F-S	Severe
52	CYS	ND4	3	C-Y	Severe
58	SER	ND4	3	S-P	Medium
70	THR	ND4	3	T-A	Mild
85	SER	ND4	3	S-F	Severe
86	SER	ND4	3	S-G	Medium
104	ILE	ND4	3	I-V	Mild
107	ILE	ND4	3	I-V	Mild
121	PHE	ND4	3	F-L	Mild
129	THR	ND4	3	T-S	Mild
165	ILE	ND4	3	I-V	Mild
166	TYR	ND4	3	Y-H	Mild
170	THR	ND4	3	T-I	Severe
187	SER	ND4	3	S-P	Medium
193	ASN	ND4	3	N-S	Medium
201	MET	ND4	3	M-T	Severe
225	ILE	ND4	3	I-V	Mild
232	ALA	ND4	3	A-T	Mild
291	ILE	ND4	3	I-V	Mild
317	ILE	ND4	3	I-V	Mild
320	GLY	ND4	3	G-E	Severe
350	THR	ND4	3	T-A	Mild
365	ALA	ND4	3	A-T	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
391	ILE	ND4	3	I-V	Mild
402	VAL	ND4	3	V-L	Mild
418	SER	ND4	3	S-L	Severe
418	SER	ND4	3	S-P	Medium
419	LEU	ND4	3	L-F	Mild
420	THR	ND4	3	T-A	Mild
421	HIS	ND4	3	H-Y	Mild
424	ASN	ND4	3	N-K	Mild
425	ASN	ND4	3	N-S	Medium
434	ASN	ND4	3	N-H	Mild
435	THR	ND4	3	T-I	Severe
445	LEU	ND4	3	L-I	Mild
455	THR	ND4	3	T-I	Severe
459	SER	ND4	3	S-F	Severe
459	SER	ND4	3	S-P	Medium
21	VAL	ND4L	3	V-L	Mild
21	VAL	ND4L	3	V-M	Mild
47	MET	ND4L	3	M-I	Mild
62	ALA	ND4L	3	A-T	Mild
62	ALA	ND4L	3	A-V	Medium
79	VAL	ND4L	3	V-I	Mild
80	SER	ND4L	3	S-A	Mild
1	MET	ND5	3	M-V	Mild
2	THR	ND5	3	T-I	Severe
2	THR	ND5	3	T-A	Mild
14	SER	ND5	3	S-A	Mild
18	PRO	ND5	3	P-L	Severe
21	THR	ND5	3	T-I	Severe
22	THR	ND5	3	T-I	Severe
22	THR	ND5	3	T-A	Mild
27	ASN	ND5	3	N-S	Medium
30	ASN	ND5	3	N-S	Medium
34	HIS	ND5	3	H-Y	Mild
38	SER	ND5	3	S-T	Mild
39	ILE	ND5	3	I-V	Mild
40	VAL	ND5	3	V-I	Mild
45	ILE	ND5	3	I-V	Mild
52	THR	ND5	3	T-C	Severe
52	THR	ND5	3	T-M	Severe

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
53	MET	ND5	3	M-V	Mild
59	GLN	ND5	3	Q-L	Severe
63	ILE	ND5	3	I-V	Mild
65	ASN	ND5	3	N-S	Medium
67	HIS	ND5	3	H-Y	Mild
70	THR	ND5	3	T-M	Severe
76	LEU	ND5	3	L-F	Mild
76	LEU	ND5	3	L-V	Mild
88	MET	ND5	3	M-T	Severe
93	ALA	ND5	3	A-T	Mild
109	ASN	ND5	3	N-S	Medium
109	ASN	ND5	3	N-Y	Medium
113	ASN	ND5	3	N-S	Medium
127	THR	ND5	3	T-A	Mild
132	VAL	ND5	3	V-I	Mild
141	PHE	ND5	3	F-L	Mild
147	VAL	ND5	3	V-M	Mild
162	ALA	ND5	3	A-T	Mild
169	ILE	ND5	3	I-T	Severe
172	ILE	ND5	3	I-V	Mild
190	ILE	ND5	3	I-M	Mild
190	ILE	ND5	3	I-V	Mild
196	TRP	ND5	3	W-L	Severe
206	ALA	ND5	3	A-T	Mild
209	SER	ND5	3	S-G	Medium
211	THR	ND5	3	T-A	Mild
270	SER	ND5	3	S-C	Severe
271	PRO	ND5	3	P-S	Medium
273	ILE	ND5	3	I-V	Mild
283	ILE	ND5	3	I-T	Severe
290	VAL	ND5	3	V-I	Mild
304	PHE	ND5	3	F-L	Mild
317	ILE	ND5	3	I-V	Mild
323	HIS	ND5	3	H-Y	Mild
331	THR	ND5	3	T-A	Mild
377	SER	ND5	3	S-T	Mild
378	LEU	ND5	3	L-Q	Severe
381	ALA	ND5	3	A-T	Mild
400	ASN	ND5	3	N-S	Medium

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
403	TYR	ND5	3	Y-D	Severe
413	LEU	ND5	3	L-I	Mild
420	SER	ND5	3	S-G	Medium
423	SER	ND5	3	S-R	Severe
424	THR	ND5	3	T-A	Mild
440	LEU	ND5	3	L-F	Mild
441	THR	ND5	3	T-I	Severe
442	ASN	ND5	3	N-S	Medium
447	ASN	ND5	3	N-D	Mild
456	ARG	ND5	3	R-G	Severe
469	THR	ND5	3	T-I	Severe
472	ILE	ND5	3	I-M	Mild
473	SER	ND5	3	S-F	Severe
473	SER	ND5	3	S-P	Medium
476	SER	ND5	3	S-A	Mild
478	PHE	ND5	3	F-I	Mild
482	ILE	ND5	3	I-T	Severe
482	ILE	ND5	3	I-M	Mild
485	TYR	ND5	3	Y-C	Severe
489	THR	ND5	3	T-M	Severe
489	THR	ND5	3	T-A	Mild
491	LEU	ND5	3	L-F	Mild
492	ALA	ND5	3	A-T	Mild
493	VAL	ND5	3	V-I	Mild
495	PHE	ND5	3	F-S	Severe
500	THR	ND5	3	T-A	Mild
508	THR	ND5	3	T-A	Mild
509	ASN	ND5	3	N-S	Medium
513	MET	ND5	3	M-T	Severe
515	SER	ND5	3	S-T	Mild
524	ASN	ND5	3	N-S	Medium
525	MET	ND5	3	M-L	Mild
531	SER	ND5	3	S-I	Severe
536	THR	ND5	3	T-M	Severe
546	GLN	ND5	3	Q-L	Severe
556	THR	ND5	3	T-I	Severe
564	LYS	ND5	3	K-M	Severe
565	THR	ND5	3	T-A	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
569	HIS	ND5	3	H-Y	Mild
571	ILE	ND5	3	I-V	Mild
575	ILE	ND5	3	I-T	Severe
575	ILE	ND5	3	I-M	Mild
576	ILE	ND5	3	I-V	Mild
577	THR	ND5	3	T-A	Mild
579	THR	ND5	3	T-A	Mild
596	ILE	ND5	3	I-L	Mild
596	ILE	ND5	3	I-V	Mild
602	ILE	ND5	3	I-M	Mild
603	THR	ND5	3	T-M	Severe
2	MET	ND6	3	M-T	Severe
4	ALA	ND6	3	A-T	Mild
34	VAL	ND6	3	V-I	Mild
35	SER	ND6	3	S-C	Severe
37	VAL	ND6	3	V-A	Medium
38	VAL	ND6	3	V-I	Mild
41	VAL	ND6	3	V-I	Mild
41	VAL	ND6	3	V-A	Medium
42	ILE	ND6	3	I-V	Mild
45	ASN	ND6	3	N-S	Medium
54	MET	ND6	3	M-V	Mild
58	ILE	ND6	3	I-T	Severe
81	ALA	ND6	3	A-T	Mild
85	GLY	ND6	3	G-E	Severe
90	VAL	ND6	3	V-A	Medium
94	VAL	ND6	3	V-A	Medium
97	ALA	ND6	3	A-P	Medium
102	LEU	ND6	3	L-F	Mild
106	VAL	ND6	3	V-A	Medium
120	SER	ND6	3	S-M	Severe
111	ALA	NDUFA11	3	A-T	Mild
128	ARG	NDUFA11	3	R-W	Severe
134	LEU	ND6	3	L-F	Mild
135	ILE	ND6	3	I-V	Mild
139	PRO	ND6	3	P-L	Severe
139	PRO	ND6	3	P-S	Medium
140	ILE	ND6	3	I-V	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
142	ALA	ND6	3	A-V	Medium
150	ARG	ND6	3	R-C	Severe
150	ARG	ND6	3	R-P	Severe
155	VAL	ND6	3	V-I	Mild
156	THR	ND6	3	T-I	Severe
156	THR	ND6	3	T-A	Mild
159	THR	ND6	3	T-A	Mild
159	THR	ND6	3	T-P	Medium
161	PHE	ND6	3	F-L	Mild
162	VAL	ND6	3	V-A	Medium
174	ASN	ND6	3	N-S	Medium
2	ARG	NDUFA10	3	R-C	Severe
137	ALA	NDUFA10	3	A-V	Medium
142	GLY	NDUFA10	3	G-R	Severe
165	PRO	NDUFA10	3	P-L	Severe
19	ARG	NDUFA11	3	R-H	Mild
65	VAL	NDUFA11	3	V-F	Medium
99	THR	NDUFA11	3	T-I	Severe
60	ARG	NDUFA12	3	R-Q	Mild
24	LEU	NDUFA13	3	L-W	Severe
41	THR	NDUFA13	3	T-I	Severe
41	THR	NDUFA13	3	T-N	Medium
80	ARG	NDUFA13	3	R-Q	Mild
87	ARG	NDUFA13	3	R-W	Severe
34	ASP	NDUFA2	3	D-N	Mild
62	PRO	NDUFA2	3	P-L	Severe
63	LYS	NDUFA2	3	K-R	Medium
93	VAL	NDUFA2	3	V-I	Mild
11	ALA	NDUFA3	3	A-G	Medium
49	PRO	NDUFA3	3	P-T	Medium
2	GLY	NDUFA5	3	G-C	Severe
63	ARG	NDUFA6	3	R-Q	Mild
4	THR	NDUFA7	3	T-P	Medium
7	ILE	NDUFA7	3	I-S	Severe
29	GLU	NDUFA7	3	E-K	Mild
37	PRO	NDUFA7	3	P-S	Medium
50	ASN	NDUFA7	3	N-S	Medium
57	ASP	NDUFA7	3	D-G	Medium
59	ARG	NDUFA7	3	R-C	Severe

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
43	MET	NDUFA8	3	M-V	Mild
54	ARG	NDUFA8	3	R-Q	Mild
134	ARG	NDUFA8	3	R-Q	Mild
21	ALA	NDUFA9	3	A-V	Medium
135	LEU	NDUFA9	3	L-M	Mild
187	ARG	NDUFA9	3	R-Q	Mild
206	TYR	NDUFA9	3	Y-C	Severe
235	ARG	NDUFA9	3	R-Q	Mild
249	ALA	NDUFA9	3	A-T	Mild
342	ILE	NDUFA9	3	I-M	Mild
33	ARG	NDUFB1	3	R-Q	Mild
38	ARG	NDUFB1	3	R-W	Severe
34	VAL	NDUFB10	3	V-L	Mild
65	PRO	NDUFB10	3	P-R	Severe
109	ARG	NDUFB10	3	R-M	Severe
164	LYS	NDUFB10	3	K-E	Mild
55	VAL	NDUFB11	3	V-I	Mild
56	TRP	NDUFB11	3	W-C	Severe
60	VAL	NDUFB3	3	V-A	Medium
2	PHE	NDUFB4	3	F-L	Mild
25	SER	NDUFB4	3	S-F	Severe
11	ILE	NDUFB5	3	I-F	Mild
87	TYR	NDUFB5	3	Y-H	Mild
137	PRO	NDUFB5	3	P-L	Severe
10	ARG	NDUFB6	3	R-Q	Mild
20	ARG	NDUFB6	3	R-Q	Mild
25	GLN	NDUFB6	3	Q-E	Mild
42	MET	NDUFB6	3	M-V	Mild
80	ILE	NDUFB6	3	I-T	Severe
108	THR	NDUFB6	3	T-I	Severe
29	GLY	NDUFB7	3	G-C	Severe
43	GLN	NDUFB7	3	Q-R	Mild
48	ALA	NDUFB7	3	A-T	Mild
55	ARG	NDUFB7	3	R-Q	Mild
61	HIS	NDUFB7	3	H-D	Medium
70	ARG	NDUFB7	3	R-C	Severe
83	ARG	NDUFB7	3	R-W	Severe
92	ARG	NDUFB7	3	R-H	Mild
105	ARG	NDUFB7	3	R-G	Severe

Residue number	Identity	Subunit	Polymorphism class	Residue cha3nge	Polymorphism change type
110	ARG	NDUFB7	3	R-Q	Mild
112	LYS	NDUFB7	3	K-Q	Mild
28	MET	NDUFB8	3	M-I	Mild
38	ASP	NDUFB8	3	D-N	Mild
75	PRO	NDUFB8	3	P-L	Severe
93	THR	NDUFB8	3	T-A	Mild
94	PRO	NDUFB8	3	P-R	Severe
3	LEU	NDUFB9	3	L-W	Severe
7	PRO	NDUFB9	3	P-L	Severe
36	TYR	NDUFB9	3	Y-H	Mild
46	ARG	NDUFB9	3	R-L	Severe
128	ARG	NDUFB9	3	R-Q	Mild
143	THR	NDUFB9	3	T-M	Severe
174	ARG	NDUFB9	3	R-Q	Mild
176	ARG	NDUFB9	3	R-Q	Mild
178	MET	NDUFB9	3	M-T	Severe
41	GLU	NDUFC1	3	E-Q	Mild
35	PHE	NDUFC2	3	F-L	Mild
59	ARG	NDUFC2	3	R-H	Mild
119	ARG	NDUFC2	3	R-H	Mild
6	SER	NDUFS1	3	S-G	Medium
122	MET	NDUFS1	3	M-L	Mild
280	THR	NDUFS1	3	T-S	Mild
331	LEU	NDUFS1	3	L-H	Severe
334	LEU	NDUFS1	3	L-F	Mild
479	MET	NDUFS1	3	M-R	Severe
515	ARG	NDUFS1	3	R-G	Severe
634	ASP	NDUFS1	3	D-N	Mild
646	ASN	NDUFS1	3	N-S	Medium
689	LYS	NDUFS1	3	K-R	Medium
80	ILE	NDUFS2	3	I-V	Mild
379	VAL	NDUFS2	3	V-L	Mild
409	HIS	NDUFS2	3	H-Y	Mild
100	ARG	NDUFS3	3	R-C	Severe
94	SER	NDUFS4	3	S-A	Mild
115	SER	NDUFS4	3	S-T	Mild
71	THR	NDUFS5	3	T-M	Severe
92	THR	NDUFS5	3	T-A	Mild
97	HIS	NDUFS5	3	H-Q	Mild

Residue number	Identity	Subunit	Polymorphism class	Residue change	Polymorphism change type
133	ASP	NDUFS7	3	D-N	Mild
134	ARG	NDUFS7	3	R-H	Mild
148	THR	NDUFS7	3	T-M	Severe
11	SER	NDUFV1	3	S-P	Medium
32	ARG	NDUFV1	3	R-S	Severe
168	GLY	NDUFV1	3	G-D	Medium
175	VAL	NDUFV1	3	V-M	Mild
339	ARG	NDUFV1	3	R-C	Severe
418	LEU	NDUFV1	3	L-M	Mild
76	PRO	NDUFV2	3	P-T	Medium
140	ILE	NDUFV2	3	I-T	Severe

Chapter 8

References

Adams, P.D., Afonie, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D. Biol. Crystallogr*. **66**(Pt 2), 213-21.

Addgene, 2018. *pLKO.1 – TRC cloning vector protocol*. Cambridge: Addgene. Available from: http://www.addgene.org/tools/protocols/plko/. [Accessed 6th July 2015].

Agip, A.A., Blaza, J.N., Bridges, H.R., Viscomi, C., Rawson, S., Muench, S.P., Hirst, J. (2018). Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol*. **25**(7), 548-56.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi. [Accessed on multiple occasions 2015-2018].

Ames, B.N. (1966). Assay of inorganic phosphate, total phosphate and phosphatases. *Method Enzmol.* **8**, 115-18.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C., Podowski, R.M., Näslund, A.K., Eriksson, A.S., Winkler, H.H., Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **393**(6707), 133-40.

Andrews, B., Carroll, J., Ding, S., Fearnley, I.M., Walker, J.E. (2013). Assembly factors for the membrane arm of human complex I. *Proc. Natl. Acad. Sci. USA* **110**(47), 18934-39.

Angerer, H., Zwicker, K., Wumaier, Z., Sokolova, L., Heide, H., Steger, M., Kaiser, S., Nübel, E., Brutschy, B., Radermacher, M., Brandt, U., Zickermann, V. (2011). A scaffold of accessory subunits links the peripheral arm and distal proton-pumping module in mitochondrial complex I. *Biochem. J.* **437**(2), 279-88.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* **44**(W1), W147-53.

Bamber, L., Harding, M., Monne, M., Slotboom, D.J., Kunji, E.R.S. (2007) The yeast mitochondrial ADP/ATP carrier functions as a monomer in mitochondrial membranes. *Proc. Natl. Acad. Sci. USA* **104**, 10830-34.

Barad, B.A., Echols, N., Wang, R.Y., Cheng, Y., DiMaio, F., Adams, P.D., Fraser, J.S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**(10), 943-46.

Baradaran, R., Berrisford, J.M., Minhas, G.S., Sazanov, L.A. (2013). Crystal structure of the entire respiratory complex I. *Nature* **494**(7438), 443-48.

Barile, M., Brizio, C., Valenti, D., De Virgilio, C., Passarella, S. (2000). The riboflavin/FAD cycle in rat liver mitochondria. *Eur. J. Biochem.* **267**(15), 4888-900.

Baughman, J.M., Perocchi, F., Girgis, H.S., Plovanich, M., Belcher-Timme, C.A., Sanacak, Y., Bao, X.R., Strittmatter, L., Goldberger, O., Bogorad, R.L., Koteliansky, V., Mootha, V.A. (2011). Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. *Nature* **476**(7360), 341-45.

Belevich, I., Bioch, D.A., Belevich, N., Wikström, M., Verkhovsky, M.I. (2007). Exploring the proton pump mechanism of cytochrome c oxidase in real time. *Proc. Natl. Acad. Sci. USA* **104**(8), 2685-90.

Bender, H-U., Almashanu, S., Steel, G., Hu, C-A., Lin, W-W., Willis, A., Pulver, A., Valle, D. (2005). Functional consequences of PRODH missense mutations. *Am. J. Hum. Genet.* **76**(3), 409-20.

Berg, J.M., Tymoczko, J.L., Stryer, L. (2002) *Biochemistry*. 5th ed. New York: W.H Freeman.

Bernsen, P.L., Gabreëls, F.J., Ruitenbreek, W., Hamburger, H.L. (1993). Treatment of complex I deficiency with riboflavin. *J. Neurol. Sci.* **188**(2), 181-87.

Berrisford, J.M., Baradaran, R., Sazanov, L.A. (2016). Structure of bacterial respiratory complex I. *Biochim. Biophys. Acta.* **1857**(7), 892-901.

Berrisford, J.M., Sazanov, L.A. (2009). Structural basis for the mechanism of respiratory complex I. *J. Biol. Chem.* **284**, 29773-83.

Birrell, J.A., Yakovlev, G., Hirst, J. (2009). Reactions of the flavin mononucleotide in complex I: a combined mechanism describes NADH oxidation coupled to the reduction of $APAD^+$, ferricyanide, or molecular oxygen. *Biochemistry* **48**(50), 12005-13.

Blaza, J.N., Vinothkumar, K.R., Hirst, J. (2018). Structure of the deactive state of mammalian respiratory complex I. *Structure* **26**(2), 312-19.

Bolton, J.A., Wood, S.A., Kennedy, D., Don, R.H., Mattick, J.S. (1999). Retinoic acid-dependent upregulation of mouse folate receptor-alpha expression in embryonic stem cells, and conservation of alternative splicing patterns. *Gene* **230**(2), 215-24.

Brandt, U. (2011). A two-state stabilization-change mechanism for proton-pumping complex I. *Biochim. Biophys. Acta.* **1807**(10), 1364-9.

Bricker, D.K., Taylor, E.B., Schell, J.C., Orsak, T., Boutron, A., Chen, Y.C., Cox, J.E., Cardon, C.M., Van Vranken, J.G., Dephoure, N., Redin, C., Boudina, S., Gygi, S.P., Brivet, M., Thummel, C.S., Rutter, J. (2012). A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, *Drosophila*, and humans. *Science* **337**(6090), 96-100.

Bridges, H.R., Bill, E., Hirst, J. (2012). Mössbauer spectroscopy on respiratory complex I: the iron-sulphur cluster ensemble in the NADH-reduced enzyme is partially oxidised. *Biochemistry* **51**(1), 149-58.

Bridges, H.R., Mohammed, K., Harbour, M.E., Hirst, J. (2017). Subunit NDUFV3 is present in two distinct isoforms in mammalian complex I. *Biochim. Biophys. Acta.* **1858**(3), 197-207.

Buchet, K., Godinot, C. (1998). Functional F1-ATPase essential in maintain growth and membrane potential of human mitochondrial DNA-depleted ρ° cells. *J. Biol. Chem.* **237**(36), 22983-9.

Carroll, J., Ding, S., Fearnley, I.M., Walker, J.E. (2013). Post translation modifications near the quinone-binding site of mammalian complex I. *J. Biol. Chem.* **288**(34), 24799-808.

Chiong, M.A., Sim, K.G., Carpenter, K., Rhead, W., Ho, G., Olsen, R.K., Christodoulou, J. (2007). Transient multiple acyl-CoA dehydrogenation deficiency in a newborn female caused my maternal riboflavin deficiency. *Mol. Genet. Metab.* **92**(1-2), 109-14.

Chovancová, E., Pavelka, A., Beneš, P., Strnad, O., Brezovský, J., Kozlíková, B., Gora, A., Šustr, V., Klvaňa, M., Medek, P., Biedermannová, L., Sochor, J., Damborský, J. (2012). CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Computational Biology* **8**(10), e1002708.

Corona, P., Antozzi, C., Carrara, F., D'Incerti, L., Lamantea, E., Tiranti, V., Zeviani, M. (2001). A novel mtDNA mutation in the ND5 subunit of complex I in two MELAS patients. *Ann. Neurol.* **49**, 106-10.

Costantini, S., Colonna, G., Facchiano, A.M. (2008). ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation* **3**(3), 137-38.

Cotelli, M.S., Viemli, V., Rimoldi, M., Rizetto, M., Castellotti, B., Bertasi, V., Todeschini, A., Gregorelli, V., Baronchelli, C., Gellera, C., Padovani, A., Filosto, M. (2012). Riboflavin-responsive multiple acyl-coA dehydrogenase deficiency with unknown genetic defect. *Neurol. Sci.* **33**(6), 1383-87.

Cotter, D.G., Schugar, R.C., Crawford, P.A. (2013). Ketone body metabolism and cardiovascular disease. *Am. J. Physiol. Heart Circ. Physiol.* **304**(8), H1060-76.

Cox, J., Mann, M. (2008). MaxQuant enables high peptide identification rates, individualised p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**(12), 1367-72.

Crichton, P.G., Lee, Y., Riprecht, J.J., Cerson, E., Thangaratnarajah, C., King, M.S., Kunji, E.R.S. (2015). Trends in thermostability provide information on the nature of substrate, inhibitor and lipid interactions with mitochondrial carriers. *J. Biol. Chem.* **290**(13), 8206-17.

Crofts, A.R., Meinhardt, S.W., Jones, K.R., Snozzi, M. (1983). The role of the quinone pool in the cyclic eletron-transfer chain of *Rhodopseudomonas sphaeroides*: a modified Q-cycle mechanism. *Biochim. Biophys. Acta*. **723**(2), 202-18.

Darrouzet, E., Issartel, J.P., Lunardi, J., Dupuis, A. (1998). The 49 kDa subunit of NADH-ubiquinone oxidoreductase (complex I) is involved in the binding of piericidin and rotenone, two quinone-related inhibitors. *FEBS Lett.* **43**(1), 34-38.

De la Campa, E.Á., Padilla, N., de la Cruz, X. (2017). Development of pathogenicity predictors specific for variants that do not comply with clinical guidelines for the use of computational evidence. *BMC Genomics* **18**(Suppl. 5), 569.

Dimauro, S., Davidzon, G. (2005). Mitochondrial DNA and disease. *Ann. Med.* **37**(3), 222-32.

Di Luca, A., Gamiz-Hernandez, A.P., Kaila, V.R.I. (2017). Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. USA* **114**, E6314-21.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet,

H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, GF., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin, Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.L. (2004). Genome evolution in yeasts. *Nature* **430**(6995), 35-44.

Dunning, C.J., McKenzie, M., Sugiana, C., Lazarou, M., Silke, J., Connelly, A., Fletcher, J.M., Kirby, D.M., Thorburn, D.R., Ryan, M.T. (2007). Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *EMBO J.* **26**(13), 3227-37.

Efremov, R.G., Sazanov, L.A. (2011). Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414-20.

Efremov, R.G., Sazanov, L.A. (2012). The coupling mechanism of respiratory complex I – a structural and evolutionary perspective. *Biochim. Biophys. Acta* **1817**(10), 1785-95.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010). Features and development of Coot. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**(Pt4), 486-501.

Fan, Y., Dickman, K.G., Zong, W.X. (2010). Akt and c-Myc differentially activate cellular metabolic programs and prime cells to bioenergetic inhibition. *J. Biol. Chem.* **285**(10), 7324-33.

Fassone, E., Rahman, S. (2012). Complex I deficiency: clinical features, biochemistry and molecular genetics. *J. Med. Genet.* **49**(9), 578-90.

Fato, R., Bergamini, C., Bortolus, M., Maniero, A.L., Leoni, S., Ohnishi, T., Lenaz, G. (2009). Differential effects of mitochondrial complex I inhibitors on the production of reactive oxygen species. *Biochim. Biophys. Acta.* **1787**(5), 384-92.

Fato, R., Estornell, E., Di Bernardo, S., Pallotti, F., Parenti Castelli, G., Lenaz, G. (1996). Steady-state kinetics of the reduction of coenzyme Q analogs by complex I (NADH:ubiquinone oxidoreductase) in bovine heart mitochondria and submitochondrial particles. *Biochemistry* **35**(8), 2705-16.

Fedor, J.G., Jones, A.Y., Di Luca, A., Kaila, V.R.I., Hirst, J. (2017). Correlating kinetic and structural data on ubiquinone binding and reduction by respiratory complex I. *Proc. Natl. Acad. Sci. USA.* **114**(48), 12737-42.

Ferguson, S.J., Nicholls, D. (2013). *Bioenergetics.* 4th ed. London: Elsevier.

Fiedorczuk, K., Letts, J.A., Deglieposti, G., Kaszuba, K., Skehel, M., Sazanov, L.A. (2016). Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **538**(7625), 406-10.

Fiedorczuk, K., Sazanov, L.A. (2018). Mammalian mitochondrial complex I structure and disease-causing mutations. *Trends Cell Bio.* **28**(10), 835-67.

Floyd, B.J., Wilkerson, E.M., Veling, M.T., Minogue, C.E., Xia, C., Beebe, E.T., Wrobel, R.L., Cho, H., Kremer, L.S., Alston, C.L., Gromek, K.A., Dolan, B.K., Ulbrich, A., Stefely, J.A., Bohl, S.L., Werner, K.M., Jochem, A., Westphall, M.S., Rensvold, W., Taylor, R.W., Prokisch, H., Kim, J.P., Coon, J.J., Pagaliarini, D.J. (2016). Mitochondrial protein interaction mapping identifies regulators of respiratory chain function. *Mol. Cell* **63**(4), 621-32.

Formosa, L.E., Mimaki, M., Frazier, A.E., McKenzie, M., Stait, T.L., Thorburn, D.R., Stroud, D.A., Ryan, M.T. (2015). Characterisation of mitochondrial FOXRED1 in the assembly of respiratory chain complex I. *Hum. Mol. Genet.* **24**(10), 2952-65.

Freidrich, T., Hofhaus, G., Ise, W., Nehls, U., Schmitz, B., Weiss. H. (1989). A small isoform of NADH:ubiquinone oxidoreductase (complex I) without mitochondrially encoded subunits is made in chloramphenicol-treated *Neurospora crassa*. *Eur. J. Biochem.* **180**(1), 173-80.

Galkin, A., Dröse, S., Brandt, U. (2006). The proton pumping stoichiometry of purified mitochondrial complex I reconstituted in proteoliposomes. *Biochim. Biophys. Acta.* **1757**, 1575-81.

Galkin, A., Meyer, B., Wittig, I., Karas, M., Schägger, H., Vinogradov, Brandt, U. (2008). Identification of the mitochondrial ND3 subuint as a structural component involved in the active/deactive enzyme transition of respiratory complex I *J. Biol. Chem.* **283**(30), 20907-13.

Gasparre, G., Porcelli, A.M., Bonora, E., Pennisi, L.F., Toller, M., Iommarini, L., Ghelli, A., Moretti, M., Betts, C.74M., Martinelli, G.N., Ceroni, A.R., Curcio, F., Carelli, V., Rugulo, M., Tallini, G., Romeo, G. (2007). Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumours. *Proc. Natl. Acad. Sci. USA* **104**(21), 9001-6.

Gledhill, J.R., Montgomery, M.G., Leslie, A.G.W., Walker, J.E. (2007). How the regulatory protein, IF1, inhibits F1-ATPase from bovine mitochondria. *PNAS* **104**(40), 15671-76.

Gregersen, N. (1985). Riboflavin-responsive defects in beta-oxidation. *J. Inherit. Metab. Dis.* **8**, 65-9.

Grgić, L. (2004). Exploring the 49 kDa subunit as part of the core of complex I (NADH:ubiquinone oxidoreductase) from *Yarrowia lipolytica*. Thesis (PhD). Goethe University, Frankfurt.

Grgić, L., Zwicker, K., Kashini-Poor, N., Kerscher, S., Brandt, U. (2004). Functional significance of conserved histidines and arginines in the 49-kDa subunit of mitochondrial complex I. *J. Biol. Chem.* **279**, 21193-99.

Guerrero-Castillo, S., Baertling, F., Kownatzki, D., Wessels, H.J., Arnold, S., Brandt, U., Nijtmans, L. (2017). The assembly pathway of mitochondrial respiratory chain complex I. *Cell Metab.* **25**(1), 128-39.

Guo, R., Zong, S., Wu, M., Gu, J., Yang, M. (2017). Architecture of human mitochondrial respiratory megacomplex $I_2III_2IV_2$. *Cell* **170**, 1247-57.

Hao, H.X., Khalimonchuk, O., Schranders, M., Dephoure, N., Bayley, J.P., Kunst, H., Devilee, P., Cremers, C.W., Schiffman, J.D., Bentz, B.G., Gygi, S.P., Winge, D.B., Kremer, H., Rutter, J. (2009). SDH5, a gene required for the flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science* **325**(5944), 1139-42.

Heide, H., Bleier, L., Steger, M., Ackermann, J., Dröse, S., Schwamb, B., Zörnig, M., Reichert, A.S., Koch, I., Wittig, I., Brandt, U. (2012). Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab*. **16**(4), 538-49.

Hellebrekers, D.M.E.I., Sallevert, S.H.E.H., Theunissen, T.E.J., Hendrickx, A.T.M., Gottschalk, R.W., Hoeijmakers, J.G.J., Habets, D.D., Bierau, J., Schoonderwoerd, R.G., Smeets, H.J.M. (2017). Novel SLC25A32 mutation in a patient with a severe neuromuscular phenotype. *Eur. J. Hum. Genet.* **25**(7), 886-88.

Herzig, S., Raemy, E., Montessuit, S., Veuthey, J.L., Zamboni, N., Westermann, B., Kunji, E.R., Martinou, J.C. (2012). Identification and functional expression of the mitochondrial pyruvate carrier. *Science* **337**(6090), 93-96.

Heuts, D.P., Scrutton, N.S., McIntire, W.S., Fraaije, M.W. (2009). What's in a covalent bond? On the role and formation of covalently bound flavin cofactors. *FEBS J.* **276**(13), 3405-27.

Hinchliffe, P., Carroll, J., Sazanov, L.A. (2006). Identification of a novel subunit of the respiratory complex I from *Thermus thermophilus*. *Biochemistry* **45**(14), 4413-20.

Hipkiss, A.R., Cartwright, S.P., Bromley, C., Gross, S.R., Bill, R.M. (2013). Carnosine: can understanding its actions on energy metabolism and protein homeostasis inform its therapeutic potential? *Chem. Cent. J.* **7**(1) 38.

Hirst, J. (2009). Towards the molecular mechanism of respiratory complex I. *Biochem. J.* **425**(2), 327-39.

Hirst, J. (2013). Mitochondrial complex I. *Annu. Rev. Biochem.* **82**, 551-75.

Hirst, J., Roessler, M.M. (2016). Energy conversion, redox catalysis and generation of reactive oxygen species by respiratory complex I. *Biochim. Biophys. Acta.* 1**857**, 872-83.

Hoefs, S.J., Roenburg, R.J., Smeitink, J.A., van den Heuvel, L.P. (2012). Molecular basis of biochemical complex I deficiency. *Mitochondrion* **12**(5), 520-32.

Höllerhage, M., Matusch, A., Champy, P., Lombès, A., Ruberg, M., Oertel, W.H., Höglinger, G.U. (2009). Natural lipophilic inhibitors of mitochondrial complex I are candidate toxins for sporadic neurodegenerative tau pathologies. *Exp. Neurol.* **220**(1), 133-42.

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Scheukter, J., Carpten, J.D>, Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittlemore, A.S., Sieh, W. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**(4), 877-85.

Irwin, M.H., Parameshwaran, K., Pinkert, C.A. (2013). Mouse models of mitochondrial complex I deficiency. *Int. J. Biochem. Cell Biol.* **45**(1), 34-40.

Iwata, S., Lee, J.W., Okada, K., Lee, J.K., Iwata, M., Rasmussen, B., Link, T.A., Ramaswamy, S., Jap, B.K. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. *Science* **281**(5373), 64-71.

Izzo, V., Bravo-San Pedro, J.M., Sica, V., Kroemer, G., Galluzzi, L. (2016). Mitochondrial permeability transition: new findings and persisting uncertainties. *Trends Cell Biol.* **26**(9), 655-67.

Jones, A.J., Blaza, J.N., Bridges, H.R., May, B., Moore, A.L., Hirst, J. (2016). A self-assembled respiratory chain that catalyses NADH oxidation by ubiquionine-10 cycling between complex I and the alterative oxidase. *Agnew. Chem. Int. Ed. Engl.* **55**(2), 728-31.

Jones, A.J., Blaza, J.N., Varghese, F., Hirst, J. (2017). Respiratory complex I in *Bos taurus* and *Paracoccus denitrificans* pumps four protons across the membrane for every NADH oxidised. *J. Biol. Chem.* **292**(12), 4987-95.

Joseph-Horne, T., Holloman, D.W., Wood, P.M. (2001). Fungal respiration: a fusion of standard and alterative oxidases. *Biochim. Biophys. Acta.* **1504**(2-3), 179-95.

Jung, C., Higgins, C.M., Xu, Z. (2000) Measuring the quantity and activity of mitochondrial electron transport chain complexes in tissues of the central nervous system using blue native polyacrylamide gel electrophoresis. *Anal. Biochem.* **286**(2), 214-23.

Kapahnke, M., Banning, A., Tikkanen, R. (2016). Random splicing of several exons caused by a single base change in the target exon of CRISPR/Cas9 mediated gene knockout. *Cells* **5**(4), pii: E45.

Kaplan, R.S., Pedersen, P.L. (1985). Determination of microgram quantities of protein in the presence of milligram levels of lipid with amido black 10B. *Anal. Biochem.* **150**(1), 97-104.

Kashani-Poor, N., Kerscher, S., Zickermann, V., Brandt, U. (2001). Efficient large scale purification of His-tagged proton translocating NADH:ubiquinone oxiodreductase (complex I) from the strictly aerobic yeast *Yarrowia lipolytica. Biochim. Biophys. Acta.* **1504**(2-3), 363-70.

Ke, B.X., Pepe, S., Grubb, D.R., Komen, J.C., Laskowski, A., Rodda, F.A., Hardman, B.M., Pitt, J.J., Ryan, M.T., Lazarou, M., Koleff, J., Cheung, M.M., Smolich, J.J., Thorburn, D.R. (2012). Tissue-specific splicing of an NDUFS6 gene-trap insertion generates a mitochondrial complex I deficiency-specific cardiomyopathy. *Proc. Natl. Acad. Sci. USA* **109**(6), 6165-70.

Kerscher, S.J., Eschemann, A., Okun, P.M., Brandt, U. (2001). External alternative NADH:ubiquinone oxiodreductase redirected to the internal face of the mitochondrial inner membrane rescues complex I deficiency in *Yarrowia lipolytica. J. Cell Sci.* **114**(Pt 21), 3915-21.

Kim, H.J., Jeong, M.Y., Na, U., Winge, D.R. (2012). Flavinylation and assembly of succinate dehydrogenase are dependent on the C-terminal tail of the flavoprotein subunit. *J. Biol. Chem.* **287**(48), 40670-79.

King, M.P., Attardi, G. (1996). Mitochondrial-mediated transformation of human ρ° cells. *Methods Enzymol.* **264**, 313-34.

King, M.S., Sharpley, M.S., Hirst, J. (2009). Reduction of hydrophilic ubiquinones by the flavin mitochondrial NADH:ubiquinone oxidoreductase (Complex I) and production of reactive oxygen species. *Biochemistry* **48**(9), 2053-62.

Koopman, W.J., Beyrath, J., Fung, C.W., Koene, S., Rodenburg, R.J., Willems, P.H., Smeitink, J.A. (2016). Mitochondrial disorders in children: toward development of small-molecule treatment strategies. *EMBO Mol. Med.* **8**(4), 311-27.

Küffner, R., Rohr, A., Schmiede, A., Krüll, C., Schulte, U. (1998). Involvement of two novel chaperones in the assembly of mitochondrial NADH:ubiquinone oxidoreductase (complex I). *J. Mol. Biol.* **283**(2), 409-17.

Kumar, S., Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *Chembiochem.* **3**(7), 604-17.

Kunji, E.R.S., Crichton, P.G. (2010). Mitochondrial carriers function as monomers. *Biochim. Biophys. Acta.* **1797**, 817-31.

Kunji, E.R.S. (2004). The role and structure of mitochondrial carriers. *FEBS Lett.* **564**, 239-44.

Kussmaul, L., Hirst, J. (2006). The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. *Proc. Natl. Acad. Sci. USA* **103**(20), 7607-12.

Lackner, L.L. (2014) Shaping the dynamic mitochondrial network. *BMC Biol.* **12**, 35.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259), 680-85.

LALIGN, 2018. Hinxton: EMBL-EBI. Available from https://www.ebi.ac.uk/Tools/psa/lalign/. [used by Dan Grba to align mouse and human complex I sequences, 2018].

Lau, A.J., Chang, T.K. (2014). Foetal bovine serum and human androstane receptor: evidence for activation of the SV23 splice variant by artemisinin, artemether, and arteether in a serum-free cell culture system. *Toxicol. Appl. Pharmacol.* **277**(2), 221-30.

Le Breton, N., Wright, J.J., Jones, A.J.Y., Salvadori, E., Bridges, H.R., Hirst, J., Roessler, M.M. 2017. Using hyperfine electron paramagnetic resonance spectroscopy to define the proton-coupled electron transfer reaction at the Fe-S cluster N2 in respiratory complex I. *J. Am. Chem. Soc.* **139**(45), 16319-26.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G, Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**(7616), 285-91. Database available from http://exac.broadinstitute.org/. [Accessed 1st June 2016].

Lenaz, G. (1998). Quinone specificity of complex I. *Biochem. Biophys. Acta.* 1364(2), 207-21.

Lenaz, G., Tiolo, G., Falasca, A.I., Genova, M.L. (2016). Complex I function in mitochondrial supercomplexes. *Biochim. Biophys. Acta.* **1857**(7), 991-1000.

Lienhart, W.D., Guipati, V., Macheroux, P. (2013). The human flavoproteome. *Arch. Biochem. Biophys.* **535**(2), 150-62.

Liolitsa, D., Rahman, S., Benton, S., Carr, L.J., Hanna, M.G. (2000). Is the mitochondrial complex I ND5 gene a hot-spot for MELAS causing mutations? *Ann. Neurol.* **53**(1), 128-32.

Lodish, H., Scott, M.P., Matsudaira, P.T., Berk, A., Kaiser, C.A., Kreiger, M., Zipursky, L., Darnell, J.E. (2008). *Molecular Cell Biology.* 6th ed. New York: W.H Freeman.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Saramady, M., Procaccio, V., Wallace, D.C. 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER. *Curr. Protoc. Bioinformatics* **44**, 1.23. 1-26. Database available from http://www.mitomap.org/. [Accessed 1st June 2016].

Love, M.I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(21), 550.

Marques, I., Duarte, M., Assunção, J., Ushakova, A.V., Videira, A. (2005). Composition of the complex I from *Neurospora crassa* and disruption of two "accessory" subunits. *Biochim. Biophys. Acta.* **1707**(2-3), 211-20.

Metsalu, T., Vilo, J. (2015). ClustVis: a web tool for visualising clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**(W1), W556-570.

Mick, D.U., Dennerlein, S., Wiese, H., Reinhold, R., Pacheu-Grau, D., Lorenzi, I., Sasarman, F., Weraarpachai, W., Shoubridge, E.A., Warscheid, B., Rehling, P. (2012). MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. *Cell* **151**(7), 1528-41.

Minczuk, M., He, J., Duch. A.M., Etterna, T.J., Chlebowski, A., Dzionek, K., Nijtmans, L.G., Huyen, M.A., Holt, I.J. (2011). TEFM (c17orf42) is necessary for transcription of human mtDNA. *Nucleic Acid Res.* **39**(10), 4284-99.

Mimaki, M., Wang, X., McKenzie, M., Thorburn, D.R., Ryan, M.T. (2012). Understanding mitochondrial complex I assembly in health and disease. *Biochim. Biophys. Acta.* **1817**(6), 851-62.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144-48.

Miyo, M., Konno, M., Colvin, H., Nishida, N., Koseki, J., Kawamoto, K., Tsunekuni, K., Nishimura, J., Hata, T., Takemasa, I., Mizushima, T., Doki, Y., Mori, M., Ishii, H. (2017). The importance of mitochondrial folate enzymes in human colorectal cancer. *Oncol. Rep.* **3**(1), 417-25.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Pigani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., Carpenter, A.E., Foo, S.Y., Stewart, S.A., Stockwell, B.R., Hacohen, N., Hahn, W.C., Lander, E.S., Sabanti, D.M., Root, D.E. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**(6), 1283-98.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42,** W401-407.

Morales-Rios, E., Montgomery, M.G., Leslie, A.G.W., Walker, J.E. (2015). Structure of ATP synthase from *Paracoccus denitrificans* determined by X-ray crystallography at 4.0 Å resolution. *Proc. Natl. Acad. Sci. USA* **112**(43), 13231-36.

Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.O., Song, C.Q., Sheel, A., Wu, Q., Ozata, D.M., Li, Y., Anderson, D.G., Emerson, C.P., Sontheimer, E.J., Moore, M.J., Weng, Z., Xue, W. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* **18**(1), 108-16.

Murai, M., Miyoshi, H. (2016). Current topics on inhibitors of respiratory complex I. *Bioichim. Biophys. Acta.* **1857**(7), 884-91.

Muramoto, K., Ohta, K., Shinzawa-Itoh, K., Kanda, K., Taniguchi, M., Nabekura, H., Yamashita, E., Tsukihara, T., Yoshikawa, S. (2010). Bovine cytochrome c oxidase structures enable O_2 reduction with minimization of reactive oxygens and provide a proton-pumping gate. *Proc. Natl. Acad. Sci. USA.* **107**(17), 7740-5.

Nagle, J.F., Tristram-Nagle, S. (2000). Structure of lipid bilayers. *Bioichim. Biophys. Acta.* **1469**(3), 159-95.

Nandigama, R.K., Edmonson, D.E. (2000). Influence of FAD structure on its binding and activity with the C460A mutant of recombinant human liver monoamine oxidase A. *J. Biol. Chem.* **275**(27), 20527-32.

Narendra, D., Tanaka, A., Suen, D.F., Youle, R.J. (2008). Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**(5), 795-803.

National Center for Biotechnology Information. (2018). *ClinVar database* [Online]. Bethesda: US National Library of Medicine. Database available from: http://www.clinvar.com/. [Accessed 1st June 2016].

National Center for Biotechnology Information. (2018). *PubChem Compound Database* [Online]. Bethesda: US National Library of Medicine. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/ [Accessed 24th April 2018].

Nelson, D.L., Cox, M.M. (2008). *Lehninger Principles of Biochemistry*. 5th ed. New York: W.H Freeman.

Nijhout, H.F., Reed, M.C., Lam, S.L., Shane, B., Gregory, J.F 3rd., Ulrich, C.M. (2006). *In silico* experimentation with a model of hepatic mitochondrial folate metabolism. *Theor. Biol. Med. Model* **3**, 40.

Nouws, J., Nijtmans, L., Houten, S.M., van den Brand, M., Huynen, M., Venselaar, H., Hoefs, S., Gloerich, J., Kronick, J., Hutchin, T., Willems, P., Rdoenburg, R., Wanders, R., van den Heuvel, L., Smeitink, J., Voegel, R.O. (2010). Acyl-coA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I. *Cell Metab.* **12**(3), 283-94.

Nunnari, J., Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell* **148**(6), 1145-59.

Ohnishi, T., Nakamaru-Ogiso, E., Ohnishi, S.T. (2010). A new hypothesis on the simultaneous direct and indirect proton pump mechanisms in NADH-quinone oxidoreductase (complex I). *FEBS Lett.* **584**, 4131-37.

Orii, K.E., Fukao, T., Song, X.Q., Mitchell, G.A., Kondo, N. (2008). Liver-specific silencing of the human gene encoding succinyl-CoA: 3-ketoacid CoA transferase. *Tohoku J. Exp. Med.* **215**(3), 227-36.

Oyedotun, K.S., Sit, C.S., Lemire, B.D. (2007). The *Saccharomyces cerevisiae* succinate dehydrogenase does not require haeme for ubiquinone reduction. *Biochim. Biophys. Acta.* **1767**(12), 1436-45.

Palade, G.E. (1952). The fine structure of mitochondria. *Anat. Rec.* **114**(3), 427-51.

Palmieri, F., Pierri, C.L. (2009). Structure and function of the mitochondrial carriers – role of the transmembrane helix P and G residues in the gating and transport mechanism. *FEBS Lett.* **584**(9), 1931-39.

Pejaver, V., Mooney, S.D., Radivjac, P. (2017). Missense variant pathogenicity predictors generalise well across a range of function-specific prediction challenges. *Hum. Mut.* **38**(9), 1092-1108.

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.J., Mort, M., Cooper, D.N., Sebat, J., Iakoucheva, L.M., Mooney, S.D., Radivojac. P. (2017). MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv* 134981. Available from: https://www.biorxiv.org/content/early/2017/05/09/134981 [Accessed 15th January 2018].

Pello, R., Martin, M.A., Carelli, V., Nijtmans, L.G., Achilli, A., Pala, M., Torroni, A., Gómez-Duràn, A., Ruiz-Pesini, E., Martinuzzi, A., Smeitink, J.A., Arenas, J., Ugalde, C. (2008). Mitochondira; DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial diseases. *Hum. Mol. Genet.* **17**(24), 4001-11.

Peralta, S., Torraco, A., Wenz, T., Garcia, S., Diaz, F., Moraes, C.T. (2014). Partial complex I deficiency due to the CNS conditional ablation of NDUFA5 results in a mild chromic encephalopathy but no increase in oxidative damage. *Hum. Mol. Genet.* **23**(6), 1399-412.

Pereira, B., Videira, A., Duarte, M. (2013). Novel insights into the role of *Neurospora crassa* NDUFAF2, an evolutionarily conserved mitochondrial complex I assembly factor. *Mol. Cell Biol.* **33**(13), 2623-34.

Petsko, G., Ringe, D. (2003). *Protein Structure and Function*. Oxford: Wiley-Blackwell.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Grennblatt, D.M., Meng, E.C., Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**(13), 1605-12.

Popp, M.W., Maquat, L.E. (2016). Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalised medicine. *Cell* **165**(6), 1319-1322.

Porter, K.R., Bonneville, M.A. (1973). *Fine structure of cells and tissues.* 4th ed. Philadelphia: Lea & Febinger.

Pryde, K.R., Hirst, J. (2011). Superoxide is produced by the reduced flavin in the mitochondrial complex I: a single, unified mechanisms that applies during both forward and reverse electron transfer. *J. Biol. Chem.* **286**(20), 18056-65

Rahman, S. (2015). Emerging aspects of treatment in mitochondrial disorders. *J. Inherit. Metab. Dis.* **38**(4), 641-53.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F. (2013). Genome engineering using the CRISPR-Cas 9 system. *Nature Protoc.* **8**(11), 2281-308.

Rhein, V.F., Carroll, J., Ding, S., Fearnley, I.M., Walker, J.E. (2015). NDUFAF7 methylates arginine 85 in the NDUFS2 subunit of human complex I. *J. Biol. Chem.* **288**(46), 33016-26.

Rhein, V.F., Carroll, J., Ding, S., Fearnley, I.M., Walker, J.E. (2016). NDUFAF5 hydroxylates NDUFS7 at an early stage in the assembly of human complex I. *J. Biol. Chem.* **219**(28), 14851-60.

Riccio, P., Aquila, H., Klingenberg, M. (1975). Purification of the carboxy-atractylate binding protein from mitochondria. *FEBS Lett.* **56**(1), 133-38.

Rich, P.R. (2017). Mitochondrial cytochrome c oxidase: catalysis, coupling and controversies. *Biochem. Soc. Trans.* **45**(3), 813-29.

Roberts, P.G., Hirst, J. (2012). The deactive form of respiratory complex I from mammalian mitochondria is a Na^+/H^+ antiporter. *J. Biol. Chem.* **287**(41), 34743-51.

Roessler, M.M., King, M.S., Robinson, A.J., Armstrong, F.A., Harmer, J., Hirst, J. (2010). Direct assignment of EPR spectra to structurally defined iron-sulphur clusters in complex I by double electron-electron resonance. *Proc. Natl. Acad. Sci. USA* **107**(5), 1930-5.

Ruitenbeek, W., Wendel, U., Hamel, B.C., Trijbels, J.M. (1996). Genetic counseling and prenatal diagnosis in disorders of the mitochondrial energy system. *J. Inherit. Metab. Dis.* **19**(4), 581-87.

Saada, A., Edvardson, S., Rapoport, M., Shaag, A., Amry, K., Miller, C., Lorberboum-Galski, H., Elpeleg, O. (2008). C6ORF66 is an assembly factor of mitochondrial complex I. *Am. J. Hum. Genet.* **82**(1), 32-38.

Saada, A., Vogel, R.O., Hoefs, S.J., van den Brand, M.A., Wessels, H.J., Willems, P.H., Venselaar, H., Shaag, A., Barghuti, F., Reish, O., Shohat, M., Huyen, M.A., Smeitink, J.A., van den Heuvel, L.P., Nijtmans, L.G. (2009). Mutations in NDUFAF3 (C3ORF60), encoding an NDUFAF4 (C6ORF66)-interacting complex I assembly protein, cause fatal neonatal mitochondrial disease. *Am. J. Hum. Genet.* **84**(6), 718-27.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., Reyes, A. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* **238**(1), 195-209.

Said, H.M., Ortiz, A., Ma, T.Y., McCloud, E. (1998). Riboflavin uptake by the human derived liver cells HepG2: mechanism and regulation. *J. Cell Physiol.* **175,** 588-94.

Said, H.M., Wang, S., Ma, T.Y. (2005). Mechanism of riboflavin uptake by cultured human retinal pigment epithelial ARPE-19 cells: possible regulation by an intracellular Ca^{2+}– calmodulin-mediated pathway. *J. Physiol.* **556,** 369-377.

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* **14**(3), 255-74.

Sazanov, L.A. (2014). The mechanism of coupling between electron transfer and proton translocation in respiratory complex I. *J. Bioenerg. Biomembr.* **46**(4), 247-53.

Sazanov, L.A., Hinchliffe, P. (2006). Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus. Science* **311**(5766), 1430-36.

Scaduto, R.C. Jnr., Grotyohann, L.W. (1999). Measurement of the mitochondrial membrane potential using fluorescent rhodamine derivatives. *Biophys. J.* **76**(1 Pt 1), 469-77.

Schägger, H., von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**(2), 223-31.

Schapira, A.H., Mann, V.M., Cooper, J.M., Dexter, D., Daniel, S.E., Jenner, P., Clark, J.B., Marsden, C.D. (1990). Anatomic and disease specificity of NADH CoQ1 reductase (complex I) deficiency in Parkison's disease. *J. Neurochem.* **55**(6), 2142-45.

Scheffler, I.E. (2015). Mitochondrial disease associated with complex I (NADH-CoQ oxidoreductase) deficiency. *J. Inherit. Metab. Dis.* **38**(3), 405-15.

Schiff, M., Veauville-Merllié, A., Su, C.H., Tzargoloff, T.E.J., Rak, M., Ogier de Baulny, H., Boutron, A., Smedts-Walters, H., Romero, N.B., Rigal, O., Rustin, P., Vianey-Saban, C., Acquaviva-Bourdain, C. (2016). SLC25A32 mutations and riboflavin-responsive exercise intolerance. *N. Engl. J. Med.* **374**(8), 795-97.

Schmidt, M., Freidrich, T., Wallrath, J., Ohnishi, T., Weiss, H. (1992). Accumulation of the pre-assembled membrane arm of NADH:ubiquinone oxidoreductase in mitochondria of manganese-limited grown *Neurospora crassa*. *FEBS Lett*. **313**(1), 8-11.

Schon, E.A., Bonilla, E., DiMauro, S. (1997) Mitochondrial DNA mutations and pathogenesis. *J. Bioenerg. Biomembr*. **29**(2), 131-49.

Sharma, V., Belevich, G., Gamiz-Hernandez, A.P., Róg, T., Vatulainen, I., Verkhovskaya, L., Wikström, M., Hummer, G., Kaila, V.R. (2015). Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. USA* **122**(37), 11571-76.

Sheftel, A.D., Stehling, O., Pierik, A.J., Netz, D.J., Kerscher, S., Elsässer, H.P., Wittig, I., Balk, J., Brandt, U., Lill, R. (2009). Human ind1, an iron sulphur cluster assembly factor for respiratory complex I. *Mol. Cell Biol.* **29**(22), 6059-73.

Sibbing, D., Pfeufer, A., Perisic, T., Mannes, A.M., Fritz-Wolf, K., Unwin, S., Sinner, M.F., Gieger, C., Gloeckner, C.J., Wichmann, H-E., Kremmer, E., Schäfer, Z., Walch, A., Hinterseer, M., Nabauer, M., Kääb, S., Kastrati, A., Schömig, A., Meitinger, T., Bornkamm, G.W., Conrad, M., von Beckerath, N. (2011). Mutations in the mitochondrial thioredoxin reductase gene TXNRD2 cause dilated cardiomyopathy. *Eur. Heart J.* **32**, 1121-33.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. (2011). Fast scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**(1), 539-47.

Sinha, P.K., Castro-Guerrero, N., Patki, G., Sato, M., Torres-Bacete, J., Sinha, S., Miyoshi, H., Matsuno-Yagi, A., Yagi, T. (2015). Conserved amino acid residues of the NuoD segment important for structure and function of *Escherichia coli* NDH-1 (complex I). *Biochemistry* **54**(3), 753-64.

Smet, J., De Paepe, B., Seneca, S., Lissens, W., Kotarsky, H., De Meirleir, L., Fellman, V., Van Coster, R. (2011). Complex III staining in blue native polyacrylamide gels. *J. Inherit Metab. Dis.* **34**(3), 741-77.

Smith, A.C., Robinson, A.J. (2016). MitoMiner v3.1, an update on the mitochondrial proteomics database. *Nucleic Acids Res.* **44**(D1), D1258-61.

Sousa, J.S., D'Imprima, E., Vonck, J. (2018). Mitochondrial respiratory chain complexes. *Subcell. Biochem.* **87**, 167-227.

Spaan, A.N., Ijlst, L., van Roermund, C.W., Wijburg, F.A., Wanders, R.J., Waterham, H.R. (2005). Identification of the human mitochondrial FAD transporter and its potential role in multiple acyl-coA dehydrogenase deficiency. *Mol. Genet. Metab.* **86**, 441-47.

Starbird, C.A., Maklashina, E., Cecchini, G., Iverson, T.M. (2015). Flavoenzymes: covalent versus noncovalent. *eLS*. John Wiley & Sons Ltd: Chichester.

Stehling, O., Lill, R. (2013). The role of mitochondria in cellular iron-sulphur protein biogenesis: mechanisms, connected processes, and diseases. *Cold Spring Harb. Perspect. Biol.* **5**(8), a011312.

Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Haywood, S., Hussain, S., Phillips, A.D., Cooper, D.N. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next generation sequencing studies. *Hum. Genet.* **136**(6), 665-77. Database available from http://www.hgmd.cf.ac.uk/ac/. [Accessed 1st June 2016].

Stewart, J.B., Chinnery, P.F. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* **16**(9), 530-42.

Stroud, D.A., Surgenor, E.E., Formosa, L.E., Reljic, B., Frazier, A.E., Dibley, M.G., Osellame, L.D., Stait, T., Beilharz, T.H., Thorburn, D.R., Salim, A., Ryan, M.T. (2016). Accessory subunits are integral for the assembly and function of human mitochondrial complex I. *Nature* **538**(7623), 123-126.

Stumpf, J.D., Saneto, R.P., Copeland, W.C. (2013). Clincial and molecular features of POLG-related mitochondrial disease. *Cold Spring Harb. Perspect. Biol.* **5**(4), a011395.

Sun, F., Huo, X., Zha, Y., Wang, A., Xu, J., Su, D., Bartlam, M., Rao, Z. (2005). Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* **121**(7), 1043-57.

Swalwell, H., Kirby, D.M., Blakely, E.L., Mitchell, A., Salemi, R., Sugiana, C., Compton, A.G., Tucker, E.J., Ke, B.X., Lamont, P.J., Turnbull, D.M., McFarland, R., Taylor, R.W., Thorburn, D.R. (2011). Respiratory chain complex I deficiency caused by mitochondrial DNA mutations. *Eur. J. Hum. Genet.* **19**(7), 769-75.

Territo, P.R., Mootha, V.K., French, S.A., Balaban, R.S. (2000). Ca^{2+} activation of heart mitochondrial oxidative phosphorylation: role of F_0F_1-ATPase. *Am. J. Physiol. Cell Physiol.* **278**(2), C423-35.

Tiffany, K.A., Roberts, D.L., Wang, M., Paschke, R., Mohsen, A-W.A., Vockley, J., Kim, J-J.P. (1997). Structure of human isovaleryl-CoA Dehydrogenase at 2.6 Å resolution: structural basis for substrate specificity. *Biochemistry* **36**, 8455-64.

Titus, S.A., Moran, R.G. (2000). Retrovirally mediated complementation of the glyB phenotype. Cloning of a human gene encoding the carrier for entry of folates into the mitochondria. *J. Biol. Chem.* **275**(47), 36811-17.

Tocilescu, M.A., Fendel, U., Zwicker, K., Dröse, S., Kerscher, S., Brandt, U. (2010). The role of a conserved tyrosine in the 49-Kda subunit of complex I for ubiquinone binding and reduction. *Biochim. Biophys. Acta.* **1797**, 625-32.

Torchetti, E.M., Brizio, C., Colella, M., Galluccio, M., Giancaspero, T.A., Indiveri, C., Roberti, M., Barile, M. (2010). Mitochondrial localisation of human FAD synthetase isoform 1. *Mitochondrion* **10,** 263-73.

TransOMIC Technologies, 2018. *CRISPR Genome Editing*. Huntsville: TransOMIC Technologies. Available from: http://www.transomic.com/Products/CRISPR-Genome-Editing.aspx. [Accessed 10th February 2018].

Trumpower, B.L. (1990). The protonmotive Q cycle. Energy transduction by coupling proton translocation to elecyron transfer by the cytochrome bc1 complex. *J. Biol. Chem.* **265**(20), 11409-12.

Tuppen, H.A.L., Hogan, V.E., He, L., Blakely, E.L., Worgan, L., Al-Dosary, Saretzki, G., Alston, C.L., Morris, A.A., Clarke, M., Jones, S., Devlin, A.M., Mansour, S., Chrzanowska-Lightowlers, Z.M.A., Thorburn, D.R., McFarland, R., Taylor, R.W. (2010). The p.M292T NDUFS2 mutation causes complex I-deficient Leigh syndrome in multiple families. *Brain* **133**(10), 2952-63.

Tyanova, S., Ternu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**(9), 731-40.

Tzagoloff, A., Jang, J., Glerum, D.M., Wu, M. (1996). FLX1 codes for a carrier protein involved in maintaining a proper balance of flavin nucleotides in the yeast mitochondria. *J. Biol. Chem.* **271**(13), 7392-97.

Uhlmann, M., Friedrich, T. (2005). EPR signals assigned to Fe/S cluster N1c of *Escherichia coli* NADH:ubiquinone oxiodreductase (complex I) derived from cluster N1a. *Biochemistry* **44**(5) 1653-8.

UniProt Consortium T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**(5), 2699.

Uziel, G., Moroni, I., Lamantea, E., Fratta, G.M., Ciceri, E., Carrara, F., Zeviani, M. (1997). Mitochondrial disease associated with the T8993G mutation of the mitochondrial ATPase 6 gene: a clinical, biochemical, and molecular study in six families. *J. Neurol. Neurosurg. Psychiatry* **63**(1), 16-22.

Varghese, F., Atcheson, E., Bridges, H.R., Hirst, J. (2015). Characterisation of clinically identified mutation in NDUFV1, the flavin-binding subunit of respiratory complex I, using a yeast model system. *Hum. Mol. Genet.* **24**(22), 6350-60.

Vartak, R.S., Semwal, M.K., Bai, Y. (2014). An update on complex I assembly: the assembly of players. *J. Bioengerg. Biomembr.* **46**(4), 323-28.

Vinothkumar, K.R., Zhu, J., Hirst, J. (2014). Architecture of mammalian respiratory complex I. *Nature* **515**(7525), 80-84.

Vogel, F., Bornhövd, C., Neupert, W., Reichert, A.S. (2006). Dynamic subcompartmentalisation of the mitochondrial inner membrane. *J. Cell Biol.* **175**(2), 237-47.

Vogel, R.O., Janssen, R.J., Ugalde, C., Grovenstein, M., Huijbens, R.J., Visch, H.J., van den Heuvel, L.P., Willems, P.H., Zeviani, M., Smeitink, J.A., Nijtmans, L.G. (2005). Human mitochondrial complex I assembly is mediated by NDUFAF1. *FEBS J.* **272**(20), 5317-26.

Vogel, R.O., Janssen, R.J., van den Brand, M.A., Dieteren, C.E., Verkaart, S., Koopman, W.J., Willems, P.H., Pluk, W., van den Heuvel, L.P., Smeitink, J.A., Nijtmans, L.G. (2007). Cytosolic signaling protein Ecsit also localises to mitochondria where it interacts with chaperone NDUFAF1 and functions in complex I assembly. *Genes Dev.* **21**(5), 615-24.

Walker, J.E., Fearnley, I.M., Gay, N.J., Gibson, B.W., Northrop, F.D., Powell, S.J., Runswick, M.J., Saraste, M., Tybulewicz, V.L. (1985). Primary structure and subunit stoichiomentry of F1-ATPase from bovine mitochondria. *J. Mol. Biol.* **14**(4), 677-701.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J 2nd, Nikoskelaninen, E.K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* **242**(4884), 1427-30.

Walters-Sen, L.C., Hashimoto, S., Thrush, D.L., Reshmi, S., Gastier-Foster, J.M., Astbury, C., Pyatt, R.E. (2015). Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol. Genet. Genomic Med.* **3**(2), 99-110.

Wang, C., Youle, R.J. (2009). The role of mitochondria in apoptosis. *Annu. Rev. Genet.* **43**, 95-118.

Watt, I.N., Montgomery, M.G., Runswick, M.J., Leslie, A.G., Walker, J.E. (2010). Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. USA* **107**(39), 16823-27.

Werner, R., Manthey, K.C., Griffin, J.B., Zempleni, J. (2005). HepG2 cells develop signs of riboflavin deficiency within four days of culture in riboflavin-deficient medium. *J. Nutr. Biochem.* **16,** 617-24.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B. 3rd., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C. (2018). MolProbity: More and better reference data for improved all-atim structure validation. *Protein Sci.* **27**(1), 293-315.

Wikström, M. (1984). Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone. *FEBS Lett.* **169**, 300-4.

Wikström M., Hummer, G. (2012). Stoichiometry of proton translocation by respiratory complex I and its mechanistic implications. *Proc. Natl. Acad. Sci. USA* **109**(12), 4431-6.

Wittig, I., Braun, H-P., Schägger, H. (2006). Blue native PAGE. *Nature Protoc.* **1**, 418-28.

Wu, M., Gu, J., Guo, R., Huang, Y., Yang, M. (2016). Structure of mammalian respiratory supercomplex $I_1III_2IV_1$. *Cell* **167**, 1598–1609.

Xia, D., Esser, L., Tang, W.K., Zhou, F., Zhou, Y., Yu, L., Yu, C.A. (2013). Structural analysis of cytochrome bc1 complexes: implications to the mechanism of function. *Biochim. Biophys. Acta.* **1827**(11-12), 1278-94.

Xia, J., Gill, E.E., Hancock, R.E. (2015). NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* **10**(6), 823-44.

Yarham, J.W., Elson, J.L., Blakely, E.L., McFarland, R., Taylor, R.W. (2010). Mitochondrial tRNA mutations and disease. *Wiley Interdiscp. Rev. RNA* **1**(2), 304-24.

Yazdanpanah, B., Wiegmann, K., Tchikov, V., Krut, O., Pongratz, C., Schramm, M., Kleinriddersm A., Wunderlich, T., Kashkar, H., Utermöhlen, O., Brüning, J.C., Schütze, S., Krönke, M. (2009). Riboflavin kinase couples TNF receptor 1 to NADPH oxidase. *Nature* **460**(7259), 1159-63.

Yonezawa, A., Inui, K. (2013). Novel riboflavin transporter family RFVT/SLC52: identification, nomenclature, functional characterisation and genetic diseases of RFVT/SLC52. *Mol. Aspects Med.* **34,** 693-701.

Yonezawa, A., Masuda, S., Katsura, T., Inui, K. (2008). Identification and functional characterisation of a novel riboflavin transporter, RFT1. *Am. J. Physiol.* **295,** 632-41.

Yoshikawa, S., Muramoto, K., Shinzawa-Itoh, K. (2011). Proton-pumping mechanism of cytochrome c oxidase. *Ann. Rev. Biophys.* **40**, 205-23.

Yu, C.A., Cen, X., Ma, H.W., Yin, Y., Yu, L., Esser, L., Xia, D. (2008). Domain conformational switch of the iron-sulphur protein in cytochrome bc1 complex is induced by the electron transfer from cytochrome b_L to b_H. *Biochim. Biophys. Acta.* **1777**(7-8), 1038-43.

Yuan, B., Latek, R., Hossbach, M., Tuschi, T., Lewitter, F. (2004). siRNA selection server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.* **32,** W130-34.

Zerbetto, E., Vergani, L., Dabbeni-Sala, F. (1997). Quantification of muscle mitochondrial oxidative phosphorylation enzymes via histochemical staining of blue native polyacrylamide gels. *Electrophoresis* **18**(1), 2059-64.

Zerbino, D.R., Achuthan, P., Akanni, W., Amod, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogo, O.G., Janacek, S.H., Juettemannm T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A,. Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Res.* **46**(D1), D754-61. Database available from http://www.ensembl.org/. [Accessed 5th September 2016].

Zhu, J., King, M.S., Yu, M., Klipcan, L., Leslie, A.G.W., Hirst, J. (2015). Structure of the subcomplex Iβ of mammalian complex I leads to new supernumerary subunit assignments. *Proc. Natl. Acad. Sci. USA* **112**(39), 12087-92.

Zhu, J., Vinothkumar, K.R., Hirst, J. (2016). Structure of mammalian respiratory complex I. *Nature* **536**(7616), 354-58.

Zickermann, V., Wirth, C., Nasiri, H., Siegmund, K., Schwalbe, H., Hunte, C., Brandt, U. (2015). Structural Biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**(6217), 44-49.

Ziegler, G.A., Vonrhein, C., Hanukoglu, I., Schulz, G.E. (1999). The structure of the adrenoxin reductase of mitochondrial P450 systems: electron transfer for steroid biosynthesis. *J. Mol. Biol.* **289**(4), 981-90.

Zurita Rendón, O., Shoubridge, E.A. (2012). Early complex I assembly defects result in rapid turnover of the ND1 subunit. *Hum. Mol. Genet.* **21**(17), 3815-24.

Zwicker, K., Galkin, A., Dröse, S., Grgic, L., Kerscher, S., Brandt, U. (2006). The redox-Bohr group associated with iron-sulphur cluster N2 of complex I. *J. Biol. Chem.* **281**, 23013-17.